TECHNISCHE UNIVERSITÄT MÜNCHEN
TUM SCHOOL OF ENGINEERING AND DESIGN

Behavioral Validity in Virtual Reality Pedestrian Simulators

Sonja Anna Elisabeth Schneider

Vollständiger Abdruck der von der TUM School of Engineering and Design der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender: Prof. Dr.-Ing. Johannes Fottner

Prüfer der Dissertation: 1. Prof. Dr. phil. Klaus Bengler

2. Prof. Joseph Kearney, Ph.D.

Die Dissertation wurde am 22.03.2021 bei der Technischen Universität München eingereicht und durch die Fakultät für Maschinenwesen am 28.09.2021 angenommen.

Acknowledgements

Although a dissertational thesis is written by an individual, many more people influence it one way or another. I cannot tell how this thesis would have turned out without the valuable advice, the critical reflections, and the emotional encouragement from my family, friends, and colleagues. I do know, however, that the past four years would have been considerably less productive, informative, and enjoyable, if it was not for them.

I would like to thank my supervisor Prof. Klaus Bengler for the opportunity to conduct this research at the Chair of Ergonomics. I greatly appreciated the freedom in pursuing my academic interests and the unique atmosphere at the institute. This unique atmosphere, of course, was created by all the wonderful colleagues, whom I would like to thank for enriching my work by their helpful comments and collaboration, but also by making me feel so comfortable.

Special thanks go to Philipp and André, whose assistance with the implementation of virtual scenarios was just as valuable as their comments on experimental design and the interpretation of results, and to Patrick, who not only suggested me to join the institute but also cleared a desk at his office.

Among all the students who assisted in this work, I would like to highlight the contributions of Madeleine Ratter and Guojin Li, whose ideas played an important role in the experiments on scenario-dependent factors. Furthermore, I would like to thank Dr. Angelika Kneidl for the helpful advice and the insights in a “different type of pedestrian simulation”. Our meetings have provided a lot of structure to my work, in particular in its early stages.

Last but not least I am grateful for the support from family and friends, not only during the time of my dissertation, but also during the studies it builds upon. Thank you for being there for me and showing interest in my work, for understanding my concerns, and, whenever necessary, for offering distraction and relaxation. Special thanks, of course, go to Jakob, who provided me not only with constant advice and encouragement, but also with the best cappuccino and cheesecake I can imagine.

Thank you.

Abstract

During recent years, virtual reality (VR) has become a key technology for investigating human behavior in both academia and industry. In traffic research, VR enables participants to experience virtual scenarios from a pedestrian's perspective. In comparison to common alternatives, this methodology facilitates the manipulation of experimental variables and thus the standardization and reproducibility of pedestrian research. In addition, it simplifies the analysis of hazardous encounters and supports the independence from physical prototypes. Despite its increasing popularity, however, a challenging question concerns the extent to which the results of simulator studies may be generalized to real-world traffic. In this context, at least two sources of potential biases must be considered: First, the use of VR technology may impact perceptual processes due to a mismatch of sensory cues, thereby altering behavioral responses in comparison to non-virtual environments. Second, common scenarios in experimental research often appear oversimplified. Hence, they may fail to replicate the demands that arise from the complexity which is characteristic for real-world traffic.

The objective of the present work is to assess the relative importance of factors that may compromise behavioral validity in VR pedestrian simulators. It starts by a theoretical overview of validation approaches in traffic research, from which practical guidelines are deduced under consideration of the individual context. This is followed by a review of state-of-the-art use cases, in which previous studies are classified according to the research objective, the technological equipment, the experimental task, and characteristics of the participant sample. A subsequent comparison of crossing decisions in two recent simulator types and on a test track provided empirical evidence on differences between the three environments. The results imply that perceptual biases may compromise the transferability to non-virtual environments, in particular with regard to the evaluation of vehicle speed. Additional differences were observed between the two simulators, indicating that hardware choices can affect the experimental outcomes. The third and final part of this thesis deals with effects of the laboratory setting and the oversimplification of scenarios. Based on a model of human information processing, a possible mismatch of real-world traffic and experimental research is identified, and potential countermeasures are discussed. Two simulator experiments served to empirically investigate this issue, focusing on intrinsic states such as motivation and distraction as well as on the complexity of traffic scenarios. While the information provided by these studies is limited to a relatively narrow range of possible mechanisms, the strongest effects were elicited by motivational incentives in the form of gamification.

In addition to providing empirical evidence on the extent to which different experimental settings produce consistent results, the present work aims to sensitize researchers to the complexity related to the concept of validity and the importance of distinguishing between its various layers. It is intended to encourage a balanced approach to pedestrian research, which accounts both for the need for experimental control and for the importance of generalizability, thereby supporting the meaningful interpretation of results in this safety-critical domain.

Zusammenfassung

Während der vergangenen Jahre hat sich Virtual Reality (VR) zu einer Schlüsseltechnologie in der Erforschung menschlichen Verhaltens entwickelt, welche sowohl im akademischen als auch im industriellen Umfeld zum Einsatz kommt. Im Kontext der Verkehrsforschung ermöglicht VR dem Nutzer, virtuelle Szenarien aus der Perspektive eines Fußgängers zu erleben. Im Vergleich zu gängigen Alternativen erleichtert diese Methodik die Kontrolle experimenteller Variablen und erhöht somit die Replizierbarkeit von Probandenstudien. Zudem ermöglicht sie die Analyse von Risikosituationen und gewährleistet eine weitgehende Unabhängigkeit von physischen Prototypen. Trotz zunehmender Popularität ist jedoch nach wie vor unklar, inwiefern sich die in Simulatorstudien gewonnenen Erkenntnisse auf den Realverkehr übertragen lassen. In diesem Zusammenhang sind wenigstens zwei Gründe für mögliche Abweichungen zu beachten: Einerseits kann der Einsatz von VR aufgrund veränderter sensorischer Voraussetzungen die menschlichen Wahrnehmungsprozesse beeinflussen und somit andere Reaktionen als nicht-virtuelle Umgebungen hervorrufen. Zum anderen erscheinen die Szenarien, die üblicherweise im Forschungskontext untersucht werden, stark vereinfacht. Es ist daher fraglich, ob sie die Anforderungen, die sich aus der für den Realverkehr charakteristischen Komplexität ergeben, angemessen abbilden.

Die Zielsetzung dieser Arbeit besteht in der Untersuchung verschiedener Faktoren, welche die Verhaltensvalidität in Fußgängersimulatoren beeinträchtigen können. Zu diesem Zweck erfolgt zunächst eine Zusammenfassung bisheriger Validierungsansätze im Rahmen der Verkehrsforschung, aus der praktische Empfehlungen unter Berücksichtigung des jeweiligen Kontextes abgeleitet werden. Daran schließt sich ein Überblick über aktuelle Anwendungsfälle an, anhand dessen eine Klassifikation bisheriger Forschung in Bezug auf die Fragestellung, technologische Aspekte, die Aufgabenstellung sowie die Eigenschaften der Stichprobe vorgenommen wird. Der nachfolgende Vergleich des Querungsverhaltens in zwei unterschiedlichen Simulatoren mit dem auf einer Teststrecke liefert Hinweise auf Unterschiede zwischen den drei Versuchsumgebungen. Diese lassen darauf schließen, dass veränderte Wahrnehmungsprozesse die Übertragbarkeit auf nicht-virtuelle Umgebungen einschränken, was insbesondere die Einschätzung der Fahrzeuggeschwindigkeit betrifft. Darüber hinaus weisen Unterschiede zwischen den beiden Simulatoren darauf hin, dass die Wahl einer bestimmten Hardwarekonfiguration die Versuchsergebnisse beeinflussen kann. Der dritte und abschließende Teil dieser Arbeit betrifft die Auswirkungen des experimentellen Kontextes sowie der mangelnden Komplexität untersuchter Szenarien. Ein Modell menschlicher Informationsverarbeitung bildet dabei die Grundlage für die Diskussion möglicher Diskrepanzen zwischen Experiment und Realverkehr sowie für die Identifikation potenzieller Gegenmaßnahmen. Eine empirische Untersuchung des Einflusses von Motivation und Ablenkung sowie der Komplexität des Verkehrsszenarios erfolgt anhand von zwei Simulatorstudien. Die deutlichsten Effekte zeigen sich dabei für motivationale Anreize in Form von Spielelementen, auch wenn in diesem Zusammenhang das eingeschränkte Spektrum betrachteter Wirkmechanismen zu beachten ist.

Neben dem empirischen Vergleich unterschiedlicher Versuchsumgebungen soll die vorliegende Arbeit die Vielschichtigkeit verdeutlichen, welche mit dem Begriff der Verhaltensvalidität einhergeht. Ziel ist eine ausgewogene Berücksichtigung der Bedeutsamkeit experimenteller Kontrolle sowie der Generalisierbarkeit der Ergebnisse. Nicht zuletzt aufgrund der Implikationen für die Verkehrssicherheit sind beide Aspekte für aussagekräftige Erkenntnisse bezüglich des Fußgängerverhaltens als gleichermaßen erforderlich anzusehen.

Table of Contents

Acronyms

APROSYS	Advanced Protection Systems (research project)
AR	Augmented Reality
CAVE	Cave Automatic Virtual Environment
CIT	Crossing Initiation Time
DK2	Development Kit 2
EU	European Union
FOV	Field of View
HMD	Head-Mounted Display
PET	Post-Encroachment Time
TTC	Time to Contact
US	United States
VR	Virtual Reality
VRU	Vulnerable Road User
WATCH-OVER	Vehicle-to-Vulnerable Road User Cooperative Communication and Sensing Technologies to Improve Transport Safety (research project)
WoZ	Wizard-of-Oz

1 Keynote on Pedestrian Research

1.1 Motives for Pedestrian Research

Pedestrians received a lot of attention during recent years. As walking is encouraged for environmental and health-related reasons, scientists and politicians search for incentives that motivate pedestrian activity (Ogilvie et al., 2007). At the same time, proceeding vehicle automation in urban areas has rendered pedestrian movements a focus of automotive engineers, who work on algorithms to reliably detect pedestrians and anticipate their imminent actions (Ahmed et al., 2019; Quintero et al., 2017).

Interest in promoting walking over motorized traffic has not emerged just recently (Atash, 1994). In fact, car use has been declining in the European Union (EU) since 2009, particularly among young and wealthy urban residents (Focas & Christidis, 2017). However, this trend is limited to some countries, whereas others, including Germany, exhibit a slow but stable growth. Furthermore, the reduction in car use temporarily coincides with the economic crisis of 2008/2009, rendering its permanence questionable. Notably, automobiles are still predominant in everyday mobility: In a recent EU-wide survey, 55% of respondents identified themselves as using a car for their most frequent trip, with percentages varying broadly between countries (Focas & Christidis, 2017).

Apart from comfort issues, potential pedestrians may be deterred by the risk of accidents. In the United States (US), the number of pedestrian fatalities rose by an alarming 35% between 2008 and 2017, whereas the total number of all other traffic fatalities declined by 6% (Retting, 2019). In the EU, pedestrian fatalities have decreased during the past decade, but their rate of decline did not match the overall downturn of 20% (European Commission, 2018). In both regions, pedestrian fatalities most commonly occurred in urban areas and at non-intersection locations (European Commission, 2018; Retting, 2019). In the US, an increase in fatalities was most pronounced during nighttime, with a startling 75% of fatalities caused by accidents after dark in 2017 (Retting, 2019). Similarly, the European Commission (2018) noted an increase in pedestrian fatalities during the winter months, in which daylight is scarce. In line with greater seasonal fluctuations, such effects were more pronounced in Nordic countries. The increased risk at darkness was already highlighted in the past, along with the request for improvements in street lighting (Grime, 1958; Smeed, 1949).

In contrast to car passengers, pedestrians do not have a solid barrier to shield them in case of an accident. Referring to traffic participants that lack such physical protection, the term vulnerable road users (VRU) highlights the need for the special consideration of pedestrians, cyclists, and motorcyclists in traffic research and planning (Otte et al., 2012). Even among other VRUs, pedestrians appear particularly exposed. The lack of a helmet renders them especially prone to head injuries (Otte et al., 2012) and they usually are, in contrast to cyclists and motorcyclists, not equipped with lighting devices.

The accident risk among pedestrians may be further influenced by their heterogeneous personal characteristics (Stoker et al., 2015). Driving a car or motorcycle requires a license and thus a minimum age as well as evidence of essential cognitive and physical abilities, whereas legal regulations on pedestrian activity are considerably more lenient (van Houten, 2011, p. 353). Children and the elderly have repeatedly been found to be overrepresented in pedestrian fatalities (SafetyNet, 2009; Stoker et al., 2015; Tight et al., 1989). Retting (2019) suggests smartphone-related distraction and alcohol consumption, both of which may be more common among pedestrians due to less restrictive legislation, as causes for rising fatalities. While the probability of alcohol impairments appears to decline with age for car drivers, it remains fairly constant for pedestrians between the US legal drinking age of 21 and the retirement age of 64 years (Retting, 2019). Additionally, the relatively unstandardized movement of pedestrians can be hard to predict, and drivers may fail to respond appropriately if they behave unexpectedly (Habibovic et al., 2013). Construction sites, parked vehicles, and the need to share infrastructure with cyclists may prevent pedestrians from staying within safe spots, thus contributing to the risk of accidents in crowded urban areas.

Although pedestrians are disproportionally at risk, action can be taken to change this. Retting (2019), for instance, reports a sharp decline in pedestrian fatalities in New York City thanks to local prevention strategies. Successful measures to elevate the safety, efficiency, and comfort of pedestrian activity will encompass a broad spectrum of approaches pertaining to education, infrastructure, and technology tailored to the particular needs of this group of road users. A turning point in the way public space is distributed may be marked by the Covid-19 pandemic: To reduce the risk of infection, walking is currently prioritized not only over private motorized vehicles, but also over public transport. At the same time, maintaining the required physical distance seems challenging on crowded sidewalks. Consequently, cities all over the world have launched initiatives to stimulate pedestrian-friendly infrastructure, hoping for the more permanent “side-effect” of enhanced public health through physical activity and the reduction of pollutants (Honey-Roses et al., 2020). Such measures should be accompanied by the systematic evaluation of pedestrian preferences and safety, in particular because potential changes in transportation patterns necessitate the re-calibration of existing behavioral models (Honey-Roses et al., 2020). Hence, although pedestrians have been a focus of urban planners and traffic researchers for decades, the interest in this particular group of road users has intensified due to recent events.

1.2 Methods for Pedestrian Research

While the term VRU was coined in the 1980s (Tight et al., 1989), the awareness that the respective group warrants particular protection has existed longer. Grime (1958, p. 152), for example, noted that “it is the pedestrians, the cyclists, and the motor-cyclists who have greatest need of protection from road accidents and from their consequences”. In the subsequent discussion of safety measures, however, the author focuses on the drivers of motorized vehicles. Such focus on the drivers’ perspective is no exception (Tight et al., 1989), and also more recent projects such as WATCH-OVER (http://www.watchover-eu.org/) and APROSYS (https://trimis.ec.europa.eu/project/advanced-protection-systems) primarily target vehicle technology rather than illuminating the patterns of pedestrian behavior that influence traffic safety. Indeed, the responsibility is often placed on the driver (van Houten, 2011, p. 353) and driver education, means of communication, and measures of passive safety are certainly relevant to the mitigation of accident risks. Nonetheless, it seems indispensable to include the pedestrians’ perspective in the development of standards that support the safety and efficiency of their movements. In this context, both the identification of risk factors and the evaluation of countermeasures rely on an appropriate methodology for collecting empirical data.

Current knowledge on pedestrian behavior builds on various sources, including both experimental and observational approaches. Except for computer simulations (e.g. Papadimitriou et al., 2009), which are themselves informed by previous measurements of pedestrian behavior, and the analysis of crash statistics (e.g. Otte et al., 2012; Pour-Rouholamin & Zhou, 2016), they typically include the recording of observable actions or subjective statements. The different settings in which such data are collected are commonly distinguished based on expected realism and the extent of experimental control (Zöller, 2015, p. 18). Such a classification is related to presumed differences between laboratory studies, which facilitate the control of confounding variables, and field experiments, which are assumed to provide high realism (for an alternative point of view see Chapter 2.1).

A more fine-grained examination may extend the concept of realism to account for both the fidelity of sensory cues and the realism of scenarios, which also relates to the awareness of individuals for the experimental context. In Figure 1, possible settings for the investigation of pedestrian behavior are therefore assigned along the following dimensions:

- **Experimental control** represents the extent to which an intended manipulation can be achieved while controlling for potential confounders. High experimental control allows to unambiguously attribute changes in the observed behavior to the influence of different experimental conditions.
- **Scenario realism** is understood as the agreement of the depicted scenarios with those encountered in real-world traffic. For instance, high realism may involve navigating to a common target in a complex and versatile environment, whereas low realism may be assumed for repetitive and unrelated actions, such as repeated street crossings between two vehicles that approach on an otherwise empty street.
- **Physical fidelity** refers to the accuracy of sensory cues. For instance, it concerns the field of view (FOV) and the visual resolution, the frequencies and amplitudes of auditory stimuli, and the options to perform relevant actions, such as crossing a street by means of actual walking. While some may understand fidelity in a broader sense, including for example psychological and functional aspects (Caird & Horrey, 2011, 5:15; Wynne et al., 2019), this term is reserved for the physically quantifiable properties of a simulation in the scope of this thesis.
- **Awareness** for the research setting refers to whether or not the involved individuals know that they are part of a scientific investigation. Such knowledge can be present for both the individual whose actions are analyzed and for the surrounding agents. Typically, high experimental control will require specific instructions and thus raise awareness for the experimental setting.

On the upper extreme of fidelity and scenario realism, we find naturalistic observations, in which neither of the agents forming the traffic scenario is aware of the ongoing experiment. Such observations may build on data from human observers, but also make use of various technological sensors. In this context, Ahmed et al. (2019) distinguish between active and passive sensors. Since the former require something to be attached to the person of interest, they seem less valuable in uncontrolled street environments. As long as the selection of scenarios and encounters is representative, such observations directly correspond to the ultimate target variable of pedestrian research: their behavior under realistic circumstances. Experimental control, in contrast, is low, since actions of neither agent can be influenced by the experimenter, but also because only naturally occurring events can be analyzed. If, for example, an observation is supposed to provide insights on the effects of infrastructural design, it may be possible to select locations that adequately represent the differences of interest. It is, however, often infeasible to equally account for weather conditions or the number, attentional state, and personality of involved individuals. Furthermore, it is impossible to observe novel technologies that are not yet available on public roads or to assess perceptual, emotional, and cognitive processes with the same preciseness as in laboratory settings. Privacy concerns may additionally affect data collection protocols.

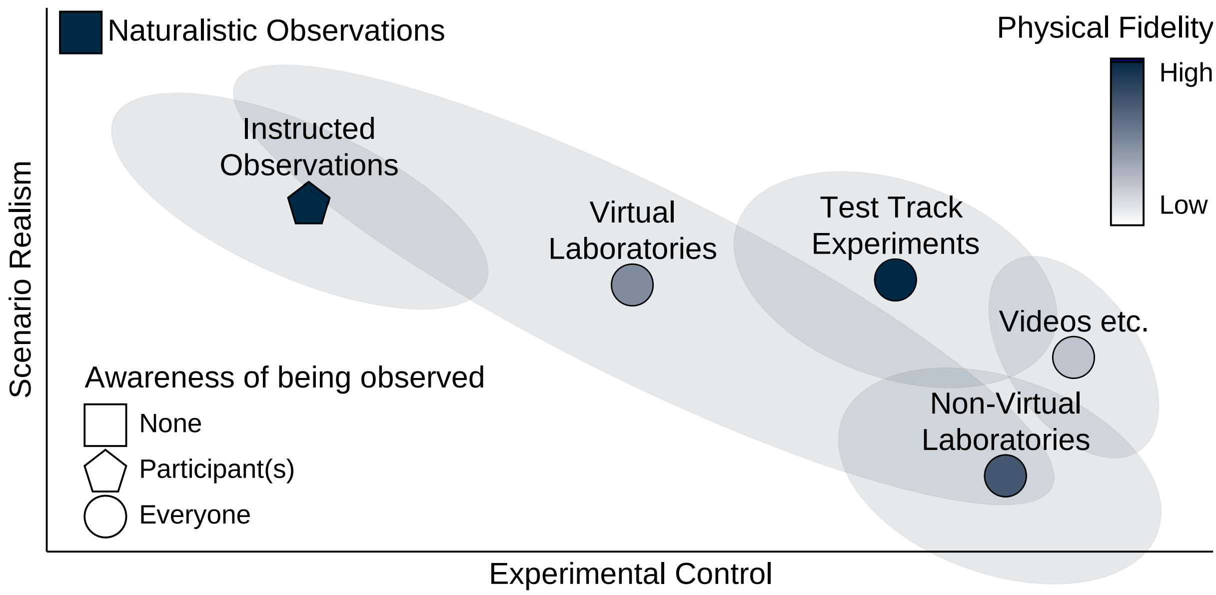

Figure 1 – Common settings for pedestrian research. The shaded areas represent a range which can reasonably be covered by the respective approach. The anchors are aligned with the middle of these regions and do not necessarily reflect the most common usage. Although the orthogonal axes imply that the dimensions are theoretically independent, the allocation of existent methods shows that experimental control and scenario realism tend to be negatively associated.

Similarly realistic scenarios may be observed in field studies, in which participants are aware of the experiment but observed within an otherwise unconstrained traffic environment (e.g. Schwebel et al., 2008). Experimental control can thereby be enhanced in comparison to naturalistic observations, for example by the purposeful selection of sample characteristics or by instructing participants to follow a certain path or search for a target. Nonetheless, results still depend on factors outside the experimenter's control, such as weather conditions and traffic density, and behavior may be biased by observer effects and social desirability. While it is possible to collect additional data on matters such as gaze behavior and subjective preferences, ethical concerns will further restrict the range of observable actions.

Experimental control is usually enhanced in laboratory settings, including both virtual and non-virtual test beds. In this case, participants are mostly aware of the experimental context, although they may not know the underlying research question. Across laboratory settings, both scenario realism and fidelity can vary. Fidelity is usually assumed to be high in advanced physical environments such as test tracks (e.g. Bellem et al., 2016), since perceptual cues like the range and amplitude of physical stimuli are equivalent to those encountered in everyday life. Physical mock-ups in indoor laboratories (e.g. Haga et al., 2015), in contrast, may provide less detailed sensory information. Furthermore, movements such as walking may be restricted due to space limitations and the risk of colliding with physical obstacles.

In virtual environments, the extent of fidelity depends on the respective setup, but is generally expected to be lower than on test tracks. This may have implications for speed, distance, and risk perception. For instance, virtual reality (VR) is known to provoke adverse effects of distance compression (Renner et al., 2013), and is often subject to additional restrictions regarding FOV and walking range. Resulting biases may induce changes in the observable behavior, which in turn can affect experimental conclusions. In terms of scenario realism, however, VR allows to simultaneously display a broad number of interacting agents in visually complex street environments and to instruct participants to perform tasks that may lead to (virtual) collisions. Limitations are imposed by restrictions in walking range, which favor a focus on repetitive crossing scenarios. Additionally, the design of complex and realistic scenarios may compromise experimental control (for a more detailed discussion, see Chapter 2.1 as well as the fourth publication included in this thesis, Schneider & Li, 2020). Physical laboratories and test tracks, in contrast, are less flexible when it comes to the number and behavior of further agents or the use of prototypes. In addition, the choice of possible tasks may be restricted to avoid hazardous situations. In summary, the understanding that test track studies are more realistic than those performed in VR may be valid with regard to the quality of sensory cues (Zöller, 2015), but needs to be reconsidered when it comes to scenario design and the evaluation of encounters that bear the risk of an accident.

Finally, pedestrian research may be complemented by low-fidelity measures such as video ratings and questionnaire studies. If traffic scenarios are described verbally, scenario realism and experimental control rely on the extent to which the description supports individual imagination. For videos, scenario realism depends on how the stimuli were generated and what sequence is depicted. It is, however, unlikely that such measures adequately reflect the affordances of multisensory cues and simultaneous events in real-world traffic. Furthermore, the range of observable measures is severely restricted, usually bound to responses that can be provided verbally or via button press. While fidelity is thus low, experimental control may be assumed relatively high as long as undisturbed viewing is guaranteed and instructions are unambiguous.

The present thesis focuses on the use of virtual environments for the investigation of pedestrian behavior. As suggested in Figure 1, this methodology comprises a particularly large range of possible configurations, whose usefulness is likely to vary for different applications. Before describing the contents of this work in more detail, some central concepts related to simulation and VR will be outlined in the following chapter.

2 Simulation and Virtual Reality in Traffic Research

Simulation implies an abstraction from a natural environment. It can be seen as a reduced replica, which contains only those elements that are considered relevant to a given task or research question (Molino et al., 2005). The challenge of creating a successful simulation thus consists in properly identifying features that can be omitted without affecting the outcomes of interest. Notably, this definition of a simulation does not specify the need for computer technology or virtual elements. In fact, the value of simulations for the analysis and modification of human behavior was recognized long before the age of ubiquitous digital computing: Flight simulators, allowing aviators to practice and demonstrate their skills in a controlled and risk-free environment, have been used since the early twentieth century (Allerton, 2009, pp. 1–3).

Nonetheless, modern simulators frequently comprise virtual environments that allow human participants to interactively explore the scenarios of interest from a first-person perspective (Schwebel et al., 2008). Interaction in this context implies that certain user inputs provoke perceivable changes in the virtual environment, such as walking movements leading to adjustments of the visual perspective. Mere videos, in contrast, may be recorded from a first-person perspective, but cannot be modified by the observer.

VR simulations offer various advantages in comparison to alternative research settings. They allow researchers to investigate potentially hazardous encounters and to represent new technologies independently from costly physical prototypes. Reproducible virtual scenarios may easily be created while precisely adjusting experimental manipulations and eliminating confounders. Additional benefits include the ease and preciseness of collecting data such as gaze behavior and the relative location of objects, the possibility to provide immediate feedback in training contexts, and the reduced emission of environmental pollutants. (de Winter et al., 2012; Zöller, 2015, pp. 19–20)

The most important disadvantage, however, stems from the uncertainty regarding the generalizability to real-world traffic. With respect to driving simulation, “several research questions may need to be answered […] related to simulator fidelity, predictive validity of driving simulators, simulator-to-reality transfer of learning, and simulator sickness” (de Winter et al., 2012, p. 48). While motion sickness appears to be less problematic for recent pedestrian simulators (Agarwal, 2019; Mallaro et al., 2017; Pala et al., 2021), the concern of validity issues has similarly been expressed in this context: “A major limitation to existing research is the lack of validation data. Before virtual reality can be properly developed as tool to understand and prevent pedestrian injuries, empirical evidence is needed to suggest pedestrian behavior in a virtual world matches behavior in the real world” (Schwebel et al., 2008, p. 1395).

The following paragraphs outline central concepts related to validity and validation theory. Importantly, precise definitions differ not only between thematic domains, but also between authors of the same discipline. The primary aim of the following classification is neither to harmonize nor to contrast their viewpoints, but to provide the terminology used throughout the remainder of this thesis.

2.1 Meanings of Validity

Originating from the idea that a valid measure provides accurate information on “what it is supposed to measure” (Newton et al., 2014, p. 18), the term validity has evolved to globally represent the agreement between a research setting and the entity it is intended to reflect. When it comes to specific use cases, however, a vast number of more or less related concepts have been proposed, fostering confusion within and across academic disciplines (Newton et al., 2014, pp. 1–3).

While more than 150 validity “types” can be found in literature (Newton et al., 2014, p. 8), only a small subset of this terminology is relevant to the present work. In the following, the focus is on the validity of using simulators in terms of a research methodology, rather than considering the measurement of a specific attribute. Accordingly, the relevant concepts are related to what Newton et al. (2014) describe as validity for research, which is required to draw scientific conclusions from empirical evidence.

Internal vs. External Validity

A traditional and very prominent distinction in experimental research refers to internal and external validity. External validity concerns the generalizability across populations, environments, and outcome variables, whereas internal validity is high if the observed changes in the dependent measure are unambiguously attributable to variations of the independent factor (Newton et al., 2014, pp. 3–4). The relative importance of both concepts likely depends on the purpose of a given experiment: According to Schram (2005), internal validity is essential when testing or modifying theories, whereas external validity seems indispensable as soon as empirical regularities shall be established. When deducing theories from laboratory results, he cautions that insufficient external validity might generate a tautology, in which theories are developed to describe the behavior in laboratories but are inappropriate for generalizing to the real world.

The relationship between internal and external validity is often considered antithetical: An increase in the similarity to real-world settings is expected to support generalizability but impair experimental control, whereas the opposite applies to more simplified, thus artificial environments (Schram, 2005). Jimenez-Buedo and Miller (2010, p. 3) describe this presumed trade-off in the following words: “the more we ensure that the treatment is isolated from potential confounds in order to ensure that the observed effect is attributable to the treatment, the more unlikely it is that the experimental results can be eloquent of phenomena of the outside world, since typically, in the outside world, many factors interact in the production of events that we are interested in.”

Leading back to the characteristics of different research settings, such artificiality may be understood as an essential difference between laboratory and field studies. As Jimenez-Buedo and Miller (2010) point out, however, artificiality may arise not only from the experimental scenario, but also from the presence of an experimenter or observer. Furthermore, they argue that “assuming that laboratory experiments are necessarily more artificial than field ones is ungranted” (p. 13), since artificiality cannot easily be quantified and likely depends on the perception of the individuals involved as well as on the measures investigated. Additionally, the concreteness of any particular real-world context in which a field experiment is conducted may restrict rather than promote the generalizability of its results (Jimenez-Buedo & Miller, 2010).

Besides the association with any particular research setting, also the general assumption of a trade-off between internal and external validity is debatable (Loudon et al., 2015). Jimenez-Buedo and Miller (2010) argue that the capacity for experimental control, but also for generalization beyond the research setting, mainly depend on the knowledge about potential confounders that is incorporated in the experimental design. The idea of a trade-off is further questioned by the understanding that if no conclusions can be drawn at all, there is no way to generalize those non-existent conclusions beyond the experimental setting. Accordingly, internal validity is often seen as a prerequisite for its external counterpart (Jimenez-Buedo & Miller, 2010). Even this unidirectional priority, however, may be questioned, given the ways in which the representativeness of stimuli and measures may affect internal validity (Persson & Wallin, 2012).

Behavioral vs. Physical Validity

A second classification, which is related in particular to the use of simulators, concerns differences between behavioral and physical validity. Physical validity is essentially synonymous to fidelity (Pinto et al., 2008), representing the agreement of physically measurable components such as vehicle dynamics and visual displays (Bellem et al., 2016). Behavioral validity, in contrast, refers to the agreement of human behavior, experience, and performance (Bellem et al., 2016). Although the latter may be further specified, for instance to contrast differences between psychological processing and observable actions (Pinto et al., 2008), I will refrain from such sub-classifications for the sake of simplicity.

When validating simulators that are meant to facilitate the investigation of human behavior, the focus should be on behavioral validity (Blaauw, 1982; Blana, 1996, p. 11; Pinto et al., 2008). Physical validity, in contrast, is mainly considered to the extent that it is expected to affect the validity of behavioral outcomes. During the past years, several studies were conducted to shed light on this relationship in both pedestrian and driving simulation, targeting divers sensory modalities such as motion cuing (Bellem et al., 2016), sound rendering (Bernhard et al., 2011), and properties related to visual perception like FOV (Goodenough, 2010), optic resolution (Jamson, 2001), and stereoscopy (Jiang et al., 2017). Although it seems reasonable to assume that behavioral validity sets minimum requirements on certain aspects of fidelity, their relationship appears to be both non-linear and contextual. The idea that "physical validity is [an] absolute necessary condition for the behavioural validity" (Blana, 1996, p. 5) may thus be replaced by the assumption that the "two aspects of validity do not have to be necessarily related" (Blana, 1996, p. 11), but frequently are so to some extent.

In addition to contradictory empirical evidence, a deterministic relationship is also improbable since a number of additional factors are thought to influence behavioral validity. Risk perception, for instance, is unlikely to match the awareness for actual danger in real-world traffic, regardless of the accuracy of sensory cues (Carsten & Jamson, 2011, p. 92; de Winter et al., 2012). Due to the complexity of this relationship and the limited knowledge about potential interactions and mediators, researchers should aim to provide a consistent level of fidelity, in which the level of detail coincides for different modes of perception. For instance, this implies the congruence of visually perceived acceleration and vestibular motion cues in driving simulation (Bellem et al., 2016). A mismatch of sensory information, which may stem from a one-sided enhancement of any particular modality, may in contrast impair behavioral validity (Bellem et al., 2016; Pinto et al., 2008; Zöller, 2015) or at least limit the effectiveness of adjustments (Carsten & Jamson, 2011, p. 91).

Absolute vs. Relative Validity

A final distinction refers to the concepts of relative and absolute validity. Relative validity can be assumed if the size and direction of effects are consistent, whereas absolute validity requires the agreement of the measured values themselves (Caird & Horrey, 2011, 5:9-5:10). Again, individual definitions may differ: Leonard and Wierwille (1975) suggest Pearson correlations as a measure of absolute validity, although a perfect correlation can be obtained even in case of a systematic offset between groups. Blana (1996, p. 18) considers relative and absolute validity as qualitative and quantitative criteria, respectively, relaxing any assumption about effect size for the former. Blaauw (1982) similarly omits this requirement and Kaptein et al. (1996) specifically understand equal effect sizes as a sign of absolute validity.

The present work entails the more conservative view that relative validity is only achieved if effect sizes are similar. Notably, all the above definitions imply that relative validity is given for any setup with absolute validity, which cannot be assumed vice versa. Furthermore, a simulator may be useful although it fails to provide absolute validity, with some authors arguing that "[f]or most research questions, establishing relative validity is sufficient" (Bellem et al., 2016, p. 443). Although absolute validity is usually necessary to deduce numerical thresholds (Bellem et al., 2016), even in such cases, a conversion factor (Molino et al., 2005) or a (possibly non-linear) arithmetic function (Mullen et al., 2011, 13:7) may be used to "translate" the results of simulator studies into real-world behavior as long as correlations between the environments are known and sufficiently high.

2.2 Simulator Validation

Validation is the process of investigating validity (Newton et al., 2014). Simulator validation, more specifically, can be achieved by "the replication of simulator and on-road tests to determine the extent to which measures correspond across contexts" (Caird & Horrey, 2011, 5:9). Since interactions between different factors may be hard to anticipate, it is usually bound to a particular scenario and research question (Bellem et al., 2016; Blana, 1996, p. 12; Kaptein et al., 1996; Pinto et al., 2008).

During the past decades, numerous studies have been performed to demonstrate the behavioral validity of driving simulators, resulting among others in multiple doctoral theses (e.g. Engen, 2008; Zöller, 2015) and review articles (e.g. Blana, 1996; Mullen et al., 2011; Wynne et al., 2019). In the more recent domain of pedestrian simulation, in contrast, empirical evidence regarding the extent of behavioral validity is less abundant. Assuming that relevant parallels exist, the following paragraphs give an overview about the findings on behavioral validity in driving simulators, before proceeding in more detail to research that focuses on pedestrian behavior. An integrative summary of validation approaches can be found in the first publication included in this thesis (Schneider & Bengler, 2020a).

Driving Simulators

An early summary of driving simulator validation since the late 1960s was provided by Blana (1996). Despite considerable technological advancements during the examined time period, studies overall seem to indicate sufficient relative but limited absolute validity (Blana, 1996, pp. 40–41). Highlighting partially conflicting results, the author concludes that, rather than high physical fidelity, "the most important element for a successful behavioural validation study is the carefully designed experimental procedure" (p. 52) and that in spite of existing data, robust answers to some of the most essential questions are missing.

Fifteen years later, Mullen et al. (2011) similarly conclude that in most cases "measures show relative validity but fail to meet requirements for absolute validity" and that one "should remain aware that simulators do not always provide an accurate picture of on-road driving behavior" (p. 13:14). In particular, behavioral variability seems to be larger in simulators than for on-road studies, and predictions based on simulator studies appear less accurate for poor drivers.

In a recent overview of 44 publications, Wynne et al. (2019) report a broad range of outcome variables considered for the purpose of simulator validation. They also note substantial differences in the understanding of "real-world" comparison data, which refer to measures as dissimilar as self-reported driving behavior, neurological assessments, the on-road driving of instrumented vehicles, and naturalistic observations via external recording devices. The authors point out that in one simulator, the agreement with real-world behavior may differ between measures, and that the validity of a given measure in turn depends on the simulator. Furthermore, behavioral validity seems to be influenced but not fully determined by simulator fidelity. Findings of lower heart rates despite increased mental effort in simulated driving, for instance, support the assumption of biased risk perception as a reasonable source of variation. (Wynne et al., 2019)

In summary, it has repeatedly been concluded that driving simulators represent a valid tool for replicating differences between experimental conditions, whereas their value for absolute predictions is limited. Additionally, researchers have demonstrated dependencies between the simulator setup, the scenario, and relevant outcome measures. Although the latter may be expected, it is noteworthy that the technological advancements of the past years do not seem to fundamentally alter this general principle. They may, however, counteract some biases and facilitate the investigation of more complex scenarios, which comprise a broader range of sensory cues and events.

Pedestrian Simulators

When comparing driving and pedestrian simulators, one must account for differences in the user's perspective and the range of common scenarios. Differences in FOV, for example, may be primarily relevant when entering a curve if driving on rural roads (Jamson, 2001), but crucial for street crossing tasks, in which cars usually approach from a 90° angle. Motion platforms, in contrast, which are a frequent concern in driving simulation (Bellem et al., 2016; Zöller, 2015), can easily be replaced by naturalistic walking for brief distances. However, the restrictions resulting from a limited tracking range thereby impose considerable limitations on the bandwidth of possible tasks. Accordingly, although most findings may be applicable on a more general level, reliable conclusions necessitate studies that specifically target pedestrian behavior.

Since modern pedestrian simulators emerged, there have been a number of attempts to empirically quantify the extent of behavioral validity in related studies. An early example was provided by Schwebel et al. (2008), who investigated street crossing behavior in 102 children and 74 adults. Comparing a screen-based setup to instructed observations in real-world traffic, they found Pearson correlations to range between .22 and .34 for the size of accepted gaps in adults, and slightly higher values of .42 and .52 for children's start delay. Furthermore, variations in crossing behavior reproduced typical age-related developments and showed reasonable associations with personality. A later analysis of the data on adult street crossing (Feldstein, 2019), however, indicated differences not only with regard to traffic density, but also in the size of accepted gaps and start delay: Smaller gaps were accepted and crossing was initiated later in the simulator. Additionally, the assumption of constant speed, which served as a proxy to calculate safety margins in the simulator, did not match on-road observations.

In a virtual scene presented via an Oculus Rift DK2 HMD, Feldstein and Dyszak (2020) instructed their participants to step backwards when they felt that an approaching vehicle had come too close to cross in front of it. In comparison to a test track setting, minimal accepted distances decreased by 26 % in VR, which matches the effects reported by Feldstein (2019). Additionally, crossing decisions in VR were strongly influenced by spatial distances, whereas participants on a physical street were more successful in accounting for the speed of vehicles.

Bhagavathula et al. (2018) compared crossing intent as well as speed and distance perception for a test track, its virtual replica, and video recordings presented via an HTC Vive HMD. Although the proportion of trials in which participants expressed an intent to cross dropped by 47.5% in VR in comparison to the test track, this difference was statistically insignificant for a sample of sixteen adults. Descriptively, however, the reluctance to accept small gaps contrasts the results of Feldstein (2019) and Feldstein and Dyszak (2020). While perceived safety did not differ between the environments, vehicle speed was perceived as higher in both the VR and the video conditions (Bhagavathula et al., 2018). Contrary to the common phenomenon of distance compression (Renner et al., 2013), distance estimates were similar on the test track and in VR (Bhagavathula et al., 2018), supporting the hypothesis that distance compression can be mitigated by technological advancements (Feldstein et al., 2020; Kelly et al., 2017). Presence was rated highest on the physical street, followed by video recordings and subsequently the VR simulation (Bhagavathula et al., 2018).

A similar approach was taken by Agarwal (2019), who analyzed differences between a test track setting, an Oculus Rift HMD, and 360° videos presented via a Samsung Gear headset. Estimated distances were found to be similar for all three environments. In contrast to Bhagavathula et al. (2018), however, also speed estimates were similar in VR and on the physical road, but more accurate for 360° videos. No significant differences were observed with regard to gap acceptance, presence, or motion sickness.

Mai (2017) compared street crossings in an HTC Vive headset to onsite observations on a university campus. Participants in the VR experiment were selected to match the demographic characteristics of those observed in real world. In contrast to the latter, participants in VR walked considerably slower (with a mean velocity of less than 0.6 m/s in comparison to 1.4 m/s on the real street). Furthermore, a majority failed to display adequate monitoring behavior before entering the street. Apparently, some participants found it difficult to follow the instructions for movement, which implied a combination of walking and teleportation mechanisms.

Fuest, Schmidt et al. (2020) investigated a pedestrian's capacity to recognize if an approaching vehicle would yield based on its acceleration profile and lateral offset. The authors compared a virtual environment presented via an HTC Vive Pro HMD to a Wizard-of-Oz (WoZ) setup. In the latter, a human driver was hidden under a seat cover to evoke the impression of an autonomous vehicle. A third and fourth condition included videos of either environment. The VR and video conditions lacked acoustic cues. In all environments, unambiguous trajectories were understood correctly, recognized faster and rated as better and less critical, whereas misinterpretations were more common for intentionally ambiguous driving. This effect was most obvious in the WoZ group. Fewer misclassifications in the remaining conditions, however, may be explained by the fact that participants took more time before making a decision, often waiting until the vehicle had either reached a complete stop or passed them. Since subjective confidence was reported only after a decision had been made, participants in the VR condition also felt more confident if a non-yielding vehicle had completed an ambiguous trajectory. Within the VR condition, participants indicated their crossing intent more than two seconds before they actually started walking. The authors conclude that although all environments proved useful to differentiate between trajectories, the behavior observed as a reaction to virtual and video scenarios is "not transferable to reality, because the pedestrians made their decisions in the WoZ setup at an earlier stage" (p. 20) and that hence "[c]onclusions as to absolute values are not possible in the VR setup" (p. 22).

Evaluating a virtual environment that was displayed via three adjoining screens, Singh et al. (2015) focused on the recognition and emotional valence of artificial sounds to facilitate the detection of electric vehicles. Although different sounds were ranked in the same order, participants in VR detected vehicles later and overall rated the sounds as less recognizable. Subjective detectability, powerfulness, and pleasantness, in contrast, did not differ significantly. Interestingly, cars were generally detected later if the time until their appearance was prolonged, which the authors understood as a sign of declining attention.

Further differences were reported for obstacle avoidance and gaze behavior. Wearing an Oculus Rift HMD with a 110° FOV, participants maintained larger lateral distances to other pedestrians than in real world (Iryo-Asano et al., 2018). In line with the findings of Mai (2017), walking speed was considerably lower. Similar results were reported by Bühler and Lamontagne (2018) for an HTC Vive HMD. When a Segway-like vehicle approached at a 90° angle, perceived danger increased in comparison to a non-virtual environment (Iryo-Asano et al., 2018). Distances were underestimated to a larger extent when facing away from the vehicle, but not for other orientations (Iryo-Asano et al., 2018). For pedestrians wearing a FOVE HMD with a 100° FOV, longer fixations, shorter saccades and a stronger focus on the center of the visual scene were observed (Berton et al., 2020). For a street view application displayed on a 17-inch monitor, the duration of fixations also differed from real world, but the direction of this effect depended on the task to perform: In line with Berton et al. (2020), fixations were longer when directed to the environment, but they were shorter when reading a map (Dong et al., 2020). Participants in the real world were considerably faster in orienting themselves and displayed different gaze patterns with regard to various types of visual stimuli. These results, however, must be seen in the light of an unequal distribution of visual information in the two environments and the limited immersion of a desktop computer (Dong et al., 2020).

In summary, empirical research on the predictive value of VR pedestrian simulators highlights some parallels to physical environments, but also demonstrates various limitations. Moreover, differences appear to apply to all stages of human behavior, including perception (Singh et al., 2015), decision-making (Bhagavathula et al., 2018; Feldstein & Dyszak, 2020), and motor responses (Iryo-Asano et al., 2018).

Importantly, all investigations are linked to a particular technological setup. Due to the short development cycles of VR equipment, also recent observations may therefore be outdated for more advanced simulators. Distance compression, for example, appears to improve in recent HMDs (Feldstein & Dyszak, 2020; Kelly et al., 2017), which frequently offer a larger FOV and extended options for naturalistic walking. Hence, options for displaying virtual scenarios will be summarized in the following paragraphs, before proceeding to the research questions that were addressed in the present thesis. A more extensive overview, including associations between different setups and common research objectives, can be found in the second publication (Schneider & Bengler, 2020b).

2.3 Simulator Technology

Head-Mounted Displays

VR based pedestrian research may cover a broad range in terms of physical fidelity. In earlier studies (e.g. Lobjois & Cavallo, 2007; Schwebel et al., 2008), the size and position of screens or projection sites and the availability of motion tracking were decisive. During the past years, however, technological advancements have produced a variety of hardware platforms, triggering extensive research on pedestrian behavior in different types of virtual environments. In particular HMDs have emerged as an economic alternative to more expensive large-screen setups such as Cave Automatic Virtual Environments (CAVE; Cavallo, Dommes et al., 2019; Jiang et al., 2017). Due to the short development cycles resulting from their widespread commercial application, ongoing technological advancements are rapidly incorporated. Additionally, versatile and easy-to-use gaming platforms can be employed to model detailed and flexible traffic scenarios.

Nonetheless, a number of restrictions apply to the investigation of pedestrian behavior in HMDs. Although shielding the user from the physical environment may support a focus on the traffic scene, the fact that participants are unable to see their physical surroundings may also influence walking patterns due to the fear of colliding with obstacles. The consciousness for the borders of the experimental room and the wired connection to a computer may prevent participants from fast movements, leading to the reduced walking speed reported previously (Iryo-Asano et al., 2018; Mai, 2017). The FOV, which is typically smaller than for natural human vision, may hinder the early detection of approaching objects and bias the perception of speeds and distances (Cavallo & Laurent, 1988; Feldstein, 2019). The accurate estimation of distances may further be complicated by the lack of a visual self-representation (Iryo-Asano et al., 2018) and the absence of alternative real-world references. Finally, the weight of an HMD may cause discomfort (Pinto et al., 2008) and contribute to biases in distance perception (Young et al., 2014).

A lot of attention has been dedicated to the underestimation of distances in comparison to non-virtual environments, which is commonly referred to as distance compression (Feldstein, 2019; Renner et al., 2013). While this phenomenon is not unique to any particular hardware (Renner et al., 2013), it apparently can be counteracted by technological advancements, as it seems less pronounced in modern HMDs (Feldstein & Dyszak, 2020; Kelly et al., 2017). In street crossing tasks, however, the appropriateness of a pedestrian's decision depends on the temporal rather than the spatial interval that remains to an approaching vehicle. According to Feldstein (2019), temporal distances are consistently underestimated, regardless of whether the estimation task is performed in a simulation, a video, or a physical environment. The author explains this by a "tendency to err in the direction of safety" (p. 777) in realistic and contextual environments. Nonetheless, humans in VR apparently find it harder to process velocities and consequently overestimate temporal intervals at higher speeds (Feldstein, 2019).

Depth perception in HMDs is further complicated as stereoscopic vision in modern headsets is commonly achieved by displaying distinct images to both eyes. This procedure leads to a conflict between accommodation and convergence, since the different images suggest larger distances than the curvature of the eyes' lenses, which are focused on the display itself (Pinto et al., 2008). Although neither accommodation nor convergence seem decisive in the range of distances in which surrounding traffic operates (Feldstein, 2019), the mismatch of information may nonetheless cause discomfort and fatigue (Mestre, 2017). Further issues may arise from the latency in refreshing the visual scene after head movements (Feldstein & Ellis, 2020; Mestre, 2017).

Alternatives

Due to their extensive use and the technological progress that can be expected in the future, the present work focuses on the application of commercially available HMDs in pedestrian research. At the same time, however, it may be worthwhile to contrast this approach to earlier desktop simulators (e.g. Schwebel et al., 2008) and more complex CAVE-like setups (Cavallo, Dommes et al., 2019). Within the group of HMDs, different technological specifications may affect the quality of sensory stimuli and the range of possible actions, thereby influencing experimental outcomes.

Young et al. (2014) compared a commercial Oculus Rift to a more elaborate HMD setup, which was commonly employed for research purposes. For comparability, the capacity for naturalistic walking was disabled in the latter. In comparison to the research device, the Oculus Rift HMD was limited in visual resolution and its potential for naturalistic movements but provided a larger FOV and was considerably more lightweight (480 g vs. 1250 g). Although it performed favorable in terms of distance perception, navigation, and visual search, motion sickness seemed more problematic, causing two participants to withdraw from the experiment.

For the purpose of road safety education, Schwebel et al. (2017) contrasted a Samsung Galaxy S6 smartphone inserted into a Cardboard viewer to a semi-immersive virtual environment consisting of three 55-inch LED monitors. Simulator sickness was rated slightly higher in the Cardboard application. Substantial correlations between both setups were found with regard to missed crossing opportunities, but not for start delay and the number of unsafe crossings. Additionally, unsafe crossings in the Cardboard were related to certain personality traits.

Despite notable differences in cost, it is unclear whether large-screen simulators such as CAVEs outperform HMDs in terms of behavioral validity. Most modern HMDs allow for walking within the range of the tracking sensors and the borders of the experimental room, whereas CAVEs further restrict walking by the size and position of their screens. Differences in the visual representation include the FOV and field of regard as well as the visibility of the user's body and the options for simultaneous use by multiple people (Mestre, 2017). When comparing a four-projection CAVE and an HTC Vive HMD with regard to crossing between continuous traffic on a single-lane street, Mallaro et al. (2017) found smaller gaps to be accepted in the HMD, whereas participants in the CAVE discriminated more strongly between gap sizes. In the HMD, participants compensated for their riskier choices by starting to cross earlier, which ultimately resulted in larger safety margins than in the CAVE condition.

Also for screen-based systems, technological specifications were shown to influence experimental outcomes. Interestingly, the quality of auditory cues appears to affect street crossing behavior (Bernhard et al., 2011), whereas the availability of stereovision seems less relevant (Jiang et al., 2017). Montuwy et al. (2017) compared buttons and a joystick as two different approaches to navigate through a virtual environment, which was displayed via three 47-inch monitors. Young adults generally preferred the joystick, whereas seniors favored the navigation via buttons. In the comparison of different screen-based setups, Maillot et al. (2017) found that the existence of differences in crossing behavior depended on the investigated age group. Practice effects related to the processing of more complex traffic scenarios, in contrast, were only evident in the more elaborate simulator.

2.4 Summary and Implications

The technological specifications of a simulator determine not only the quality of sensory cues, but also the actions that can be performed by the user. Obvious restrictions in VR pedestrian simulators include deficits in FOV and the visual resolution, the latency related to fast head movements, the absence of a visual self-representation in HMDs, and limited walking space, which restricts the range of observable scenarios. How these deficits, however, translate to changes in the experimental outcomes, is yet unclear.

In both driving and pedestrian simulation, validation is frequently understood as a comparison of environments that are perceived as physically valid to a varying degree. However, as pointed out previously, a behaviorally valid simulation requires not only sufficient sensory accuracy, but also the careful consideration of relevant features of the experimental task and scenario, since they may equally affect cognitive demands and human information processing. A comprehensive approach to evaluating VR as a tool for pedestrian research should thus include an assessment of the affordances of common scenarios and their congruence with real-world traffic.

3 Objectives and Research Questions

VR simulations have become increasingly popular in the research of pedestrian behavior. While providing a risk-free and convenient way to collect empirical data, however, the extent to which these data correspond to real-world traffic is uncertain. Besides technological restrictions, the behavioral validity in simulator studies may be compromised by the experimental setting, artificial tasks, and unrepresentative scenarios. In the present work, differences between simulated and real environments are evaluated on various levels. The aim is to contribute to the knowledge about potential biases and restrictions of VR simulators and thereby support the meaningful employment of this novel and versatile methodology.

In Chapter 1.2, physical fidelity and scenario realism were introduced as two dimensions on which research settings can be located. The third dimension, experimental control, is linked to internal validity, which mostly depends on the knowledge about confounding variables and the options to adequately consider them in data collection and analyses. Theoretically, high internal validity can also be achieved in naturalistic observations, as long as all relevant influences are known, their expression can be identified from the collected materials, and the amount of data is sufficient to adequately distinguish between various levels. Virtual environments, however, facilitate experimental control by the exact replication of arbitrary scenarios. Experimental control therefore represents a specific strength of VR experiments, whereas they are likely inferior to alternative settings in terms of physical fidelity. The role of scenario realism is discussed in more detail in the fourth publication (Schneider & Li, 2020).

Figure 2 outlines the conceptual understanding of simulator requirements in the scope of this thesis. It is based on the assumption that experiments are conducted in order to reliably and accurately predict real-world pedestrian behavior. Meaningful simulator studies, which offer a reasonable degree of experimental control based on the knowledge of potential confounders, rest on the two pillars of sufficiently accurate sensory cues and a representative range of tasks and scenarios. Since both pillars are equally important to the generalizability of the observed behavior, they must be equally considered in the scope of simulator validation. The purposeful design of experiments furthermore relies on a thorough understanding of relevant use cases and outcome variables and, lastly, applicable validation techniques.

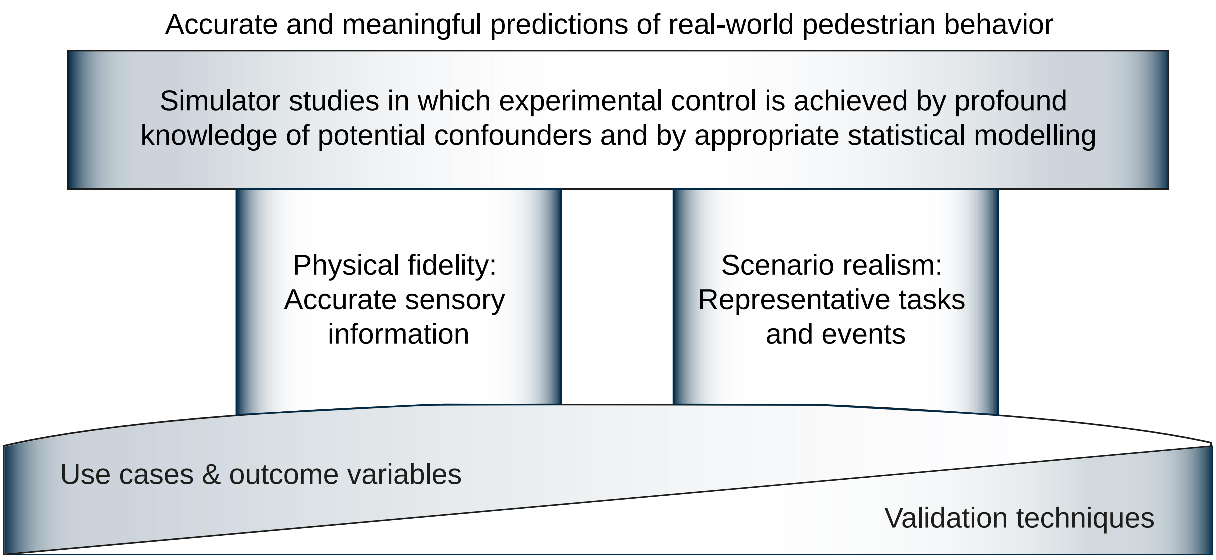

Figure 2 – Conceptual understanding of the components required to obtain reliable and meaningful conclusions from simulator studies.

The objectives of the present thesis align with these considerations. The intention is to contribute to the validation of VR pedestrian simulators by a systematic assessment of how both technological and contextual factors influence the experimental outcomes. The empiric evaluation of these aspects builds on a theoretical framework, which integrates common applications and validation techniques.

Simulator validation is related to two general questions. On the one hand, researchers aim to assess the present agreement with real-world traffic, for example by comparing a range of behavioral indicators across similar scenarios. On the other hand, validation is frequently linked to the extent to which variations in simulator properties affect human perception, decision-making, and actions. Due to the brief development cycles of VR technology, state-of-the-art assessments may suffer from a relatively short lifespan. Furthermore, generalizing across simulators is increasingly delicate, as the flexibility of current software solutions allows researchers to create very different scenarios, which are adjusted to their topic of interest, participant sample, and hardware capacities. Rather than quantifying the extent of agreement for an arbitrary setup, which may already be outdated by the time of publication, I therefore aim to assess the impact of different mechanisms that may be responsible for deviations.

Due to the broad range of parameters that presumably influence human behavior in VR, a selection must be made among possible aspects to consider. To this objective, I summarize recurring methodological characteristics and central questions related to pedestrian research and the validation of VR technology. The corresponding literature survey produces a classification of pedestrian simulator studies as well as general guidelines for simulator validation.

In this context, HMDs and CAVEs are identified as two common simulator types, which are generally expected to provide a high level of fidelity. To evaluate if pedestrian behavior is influenced by the use of VR technology, these two simulator types are compared to a non-virtual test track environment. The holistic consideration of simulator classes is thereby prioritized over the contrasting of changes in isolated technological parameters, whose impact may be moderated by other factors in any alternative setup.

With regard to scenario design, the review of recent literature reveals that current experiments mostly lack the complexity associated with real-world traffic. Consequently, an additional focus is placed on changes in pedestrian behavior due to an unequal cognitive or motivational state. The latter concerns intrinsic aspects, such as motivation and attentional focus, as well as features of the traffic scenario that raise the overall level of complexity.

In summary, the following research questions are addressed:

- What approaches do exist to evaluate behavioral validity in VR-based pedestrian research? What are the advantages and limitations of different techniques? (Schneider & Bengler, 2020a)
- How can VR-based pedestrian research be classified according to methodological and contextual aspects? (Schneider & Bengler, 2020b)
- Does pedestrian behavior in VR experiments differ from observations on test tracks? If so, is there an effect of different hardware platforms? (Schneider et al., 2021)
- To what extent is pedestrian behavior in VR experiments affected by contextual factors such as motivational state, distraction, and scenario complexity? (Schneider et al., 2019; Schneider & Li, 2020)

4 Methodology

The present work combines several methodological elements. While an initial literature survey serves as a foundation for developing a theoretical framework, resulting research questions are addressed by manipulating the factors of interest in experimental user studies. To facilitate the interpretation of the results, naturalistic observations are summarized with regard to the measures of interest in Chapter 5 and contrasted to the experimental data in Chapter 7.

4.1 Central Scenario and Dependent Measures

As a result of the literature survey outlined in the second publication (Schneider & Bengler, 2020b), street crossing is identified as the most common scenario. Notably, this includes actual walking movements as well as proxies such as stepping forward or verbalizing the intent to cross. In this context, human information processing is analyzed on different levels: Regarding perception, one may consider perceived realism (Schwebel et al., 2008), estimation accuracy (Banton et al., 2005), gaze behavior, and head rotations (Gitelman et al., 2016). Aspects of decision making refer to the type, quality, and timing of a decision (Asaithambi et al., 2016; Lobjois & Cavallo, 2009). If crossings are actually performed, one may additionally evaluate crossing patterns (Asaithambi et al., 2016), walking speed (Lobjois & Cavallo, 2009; Onelcin & Alver, 2017), acceleration, and evasive actions (Tageldin et al., 2017).

To account for its widespread application and the diversity of related measures, street crossing represents the central scenario in this thesis. Experimental studies mostly refer to gap acceptance tasks, in which participants are instructed to decide whether or not the temporal gap between approaching vehicles is sufficient to cross. In addition to the dichotomous measure of gap acceptance, temporal and spatial intervals are considered as well as velocities. In summary, the following aspects of pedestrian behavior serve as dependent measures, with subsets of them included in each of the experiments:

- **Gap acceptance** represents a binary decision of whether or not a gap of a given size is deemed sufficient to cross. This measure exclusively refers to scenarios in which cars do not yield to the pedestrian.
- The **size of accepted gaps** can be considered an alternative to the binary measure of gap acceptance. If gaps are designed to increase over time, it indicates not only the proportion of accepted trials, but also the willingness to wait for an appropriate crossing opportunity.
- **Walking speed** is calculated by dividing the distance that participants walk by the time needed to do so. The walked distance may only refer to movements that are perpendicular to the street, or additionally encompass lateral translations.
- **Crossing initiation time (CIT)** describes the time until pedestrians initiate crossing after the rear of the preceding vehicle has passed their crossing line. Negative values indicate that crossing is initiated before a vehicle has passed completely.
- **Post-encroachment time (PET)** represents a temporal safety margin referring to the intersection between a pedestrian's crossing line and the part of the street that will subsequently be occupied by the following vehicle.
- **Subjective measures** are selected according to the respective research question. They encompass self-reports of the complexity and predictability of scenarios, perceived risk, realism, and the correspondence to real-world situations, as well as questionnaires on presence (UQO Cyberpsychology Lab, 2004) and play experience (Pavlas et al., 2012).

4.2 Apparatus and Research Environments

In view of their increasing popularity, the present work specifically focuses on the behavioral validity of experimental settings in which pedestrian behavior is investigated by the use of commercial HMDs. Virtual environments are displayed via an HTC Vive Pro headset, providing a resolution of 1440 x 1600 pixels per eye at a refresh rate of 90 Hz and a nominal field of view of 110°. Auditory cues are presented via integrated headphones. All scenarios were created in the Unity games engine (Unity Technologies, 2018).

In studies that target effects of scenario design (Schneider et al., 2019; Schneider & Li, 2020), parts of the street scenery were modelled in the open source software Blender (Blender Foundation, 2017) to resemble the Munich city area (Figure 3). The simulation was run on a Hyrican Elegance 5701 PC equipped with an NVIDIA GeForce GTX 1080 Ti GPU. The participants' movements were tracked by two HTC Vive Base Stations, which were located at a height of approximately two meters in opposite corners of the experimental laboratory. Movement was thereby restricted to an area of approximately 4.5 x 4.5 meters. When crossing the street, participants walked along the diagonal of the room, which resulted in a maximum distance of 6.3 m. To account for the additional space that was needed on both crosswalks, scenarios in this setting were restricted to single-lane crossings.

Figure 3 – Sample scenes of the virtual environment created for the study presented by Schneider and Li (2020).

In the experiment outlined in the third publication (Schneider et al., 2021), the HTC Vive Pro HMD is compared to a CAVE-like simulator and a test track environment. The CAVE-like setup corresponds to the simulator described by Cavallo, Dommes et al. (2019). In this simulator, ten projection modules of 1.88 x 2.55 m^2 each form a corridor, through which participants can walk up to 7 m. At a resolution of 1400 x 1050 pixels, visuals are updated based on the position and rotation of the participant's head. The latter are tracked by eight Vicon Bonita cameras, which detect markers attached to a helmet. While spatial sound rendering is achieved by ten speakers located behind the projection modules, this setup does not provide stereovision.

For the purpose of this study, the HMD simulator was installed in the same physical space as the CAVE. An ego-avatar was visualized by applying inverse kinematics to the position of the participant's head, feet, hands, and hip. Furthermore, an HTC Vive Wireless Adapter served to avoid the restrictions that a cable would put on walking and rotation. The simulation was run on a Dell Alienware Aurora R7 PC equipped with an NVIDIA GeForce GTX 1080 Ti GPU. In both simulators, the same virtual environment was displayed, which was modelled after the test track environment. The latter corresponded to a long, single-lane street on a parking lot, which is located on the campus of the Technical University of Munich at Garching.

5 Previous Traffic Observations

A direct comparison of experimental data to naturalistic observations is complicated by a lack of experimental control in the latter. This is especially true for the fine-grained use case of street crossing, in which details such as the precise size of intervehicular gaps are decisive. Even if participants are instructed to cross in a specific location, such details cannot be controlled, and even if the matching of experimental and non-experimental data is restricted to the most crucial parameters, a substantial number of observations is needed to achieve sufficient congruence. Nonetheless, simulator studies are usually conducted to provide information and facilitate predictions with regard to real-world pedestrian behavior. Hence, it seems reasonable to consider previous findings regarding this target variable when reflecting on the extent to which experimental data approach it.

Shirazi and Morris (2015) summarize methods to collect data in real-world traffic. In addition to surveys and human observers, they name video recordings, GPS, and radar or lidar technology. For pedestrian behavior at urban intersections, they include walking speed, waiting time, route choice, and gap acceptance in their list of relevant measures. Hence, data from uninstructed observations can be obtained for most of the variables that are analyzed within the scope of this thesis, with the primary exception of subjective impressions.

At the same time, however, the way in which values of interest are defined differs. Due to changes in vehicle speed, a focus may be on the gap that remains in the moment of entering the road, rather than the intervehicular distance or PET (Asaithambi et al., 2016; Kuttan et al., 2016). Unlike CIT, waiting times often include the time for previous vehicles to pass (Asaithambi et al., 2016; Das et al., 2005). In addition, a number of contextual factors have been reported to alter pedestrians' decisions and behavior, including the presence of other individuals and parked cars, distraction, traffic density, the type and speed of approaching vehicles, and infrastructural elements such as the number and width of single lanes (Asaithambi et al., 2016; Chandra et al., 2014; Das et al., 2005; Hatfield & Murphy, 2007; Kuttan et al., 2016; Yannis et al., 2013). Finally, it remains questionable whether conclusions can be generalized across cultural and geographic regions (Pelé et al., 2017; Sueur et al., 2013). Obvious differences in contextual features therefore limit the comparability to the experimental data that are reported in the scope of this work. Nonetheless, previous observations of pedestrian behavior in on-road traffic may provide some background against which differences between experimental conditions can be judged.

5.1 Gap Acceptance

Due to their relevance for accidents, analyses of gap acceptance are relatively common. Data, however, are mostly available from countries like India, in which pedestrian traffic and fatalities account for a relatively high share of all road users (Pawar & Patil, 2015). Hence, crucial factors such as risk-taking, infrastructure, and the enforcement of traffic regulations may differ from Western Europe, which must be accounted for in the comparison to the present work.

Pawar and Patil (2015) observed four-lane divided highways in India, with two lanes of 4.5 m width in each direction. Vehicles drove at a median velocity of approximately 60 km/h. Although a zebra crossing was present, the authors state that yielding in such locations would be unusual for most drivers. Building on their observational data, they predict median accepted gaps of 4.33 s, with the 85th percentile corresponding to 5.30 s. While groups accepted smaller gaps than individuals and vehicle size was negatively related to acceptance probability, almost all gaps were accepted if they exceeded 7 s.

Although their data encompassed similar observation sites, Chandra et al. (2014) reported considerably larger gaps ranging between 6.38 and 12.78 s for a total of 17 locations in five Indian cities. In this case, however, a gap was defined as the temporal distance that remained between the subsequent car and the pedestrian at crossing initiation, regardless of whether or not there were further vehicles in between. In addition, the authors reported mean rather than median values, which may be biased by particularly large gaps in a skewed distribution.

For signalized intersections in New Delhi with a crossing width of 10 to 15 m, Das et al. (2005) predicted median accepted gaps between 3.67 and 5.75 s. Smaller estimates were obtained for narrower intersections and when pedestrians started from the median instead of the pedestrian walk. For an unmarked intersection of 15 m width, Asaithambi et al. (2016) reported mean accepted gaps of 4.21 s at an average crossing speed of 1.37 m/s. The installation of a pedestrian signal reduced these values to 3.50 s and 1.05 m/s, respectively. Kuttan et al. (2016), in contrast, observed pedestrians to accept a mean distance to the closest vehicle of 10.52 s and 8.72 s for undivided, bidirectional traffic at a street width of 8.5 and 7 m and an average crossing speed of 0.96 and 1.26 m/s, respectively. In this case, the less standardized and thus less predictable approach of vehicles in the selected locations may have resulted in more cautious behavior.

For an unsignalized single-lane street in Athens, Greece, Yannis et al. (2013) found that almost all pedestrians would start crossing if the temporal distance to the next vehicle exceeded five to six seconds. The mean accepted interval corresponded to 3.29 s at a standard deviation of 1.76 s, and a vehicle speed of 25.21 ± 7.82 km/h. Reporting similar values, Jiang et al. (2011) found that jaywalking pedestrians would usually cross in front of a vehicle at a median gap of 3.52 s in the cities of Changchun and Beijing, China. The mean value corresponded to 4.10 s, underlining differences between the two measures. The values reported by different publications are summarized in Figure 4.

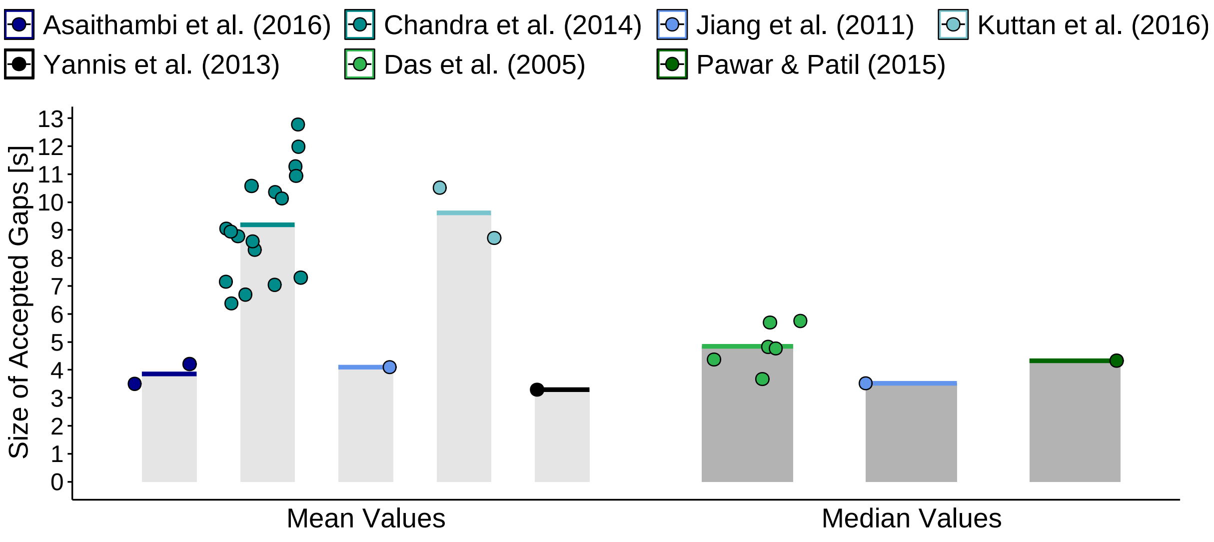

Figure 4 – Size of accepted gaps in naturalistic traffic according to different publications. Bars represent the median value for the respective publication. Values on the left refer to reports of the mean accepted gap size, whereas those on the right refer to the median.

5.2 Crossing Speed

In addition to gap acceptance, crossing speed is frequently reported due to its relevance for street design. Hatfield and Murphy (2007), for instance, analyzed the walking speed at signalized and unsignalized crossings in three suburbs of Sydney, Australia. Distinguishing between male and female pedestrians, they report mean values between approximately 1.34 and 1.86 m/s for undistracted individuals. Pelé et al. (2017), in contrast, observed a somewhat lower average speed ranging between 0.96 and 1.15 m/s for a total of seven signalized sites located in Strasbourg, France, and Nagoya, Japan.

In Israel, Gitelman et al. (2017) contrasted different options for the design of crossing infrastructure in four-lane divided roads, on which the mean vehicle speed ranged from 36.1 to 55.5 km/h, with a standard deviation of 4.5 to 11.7 km/h. The time needed to cross a width of approximately 7 meters ranged between 5.0 and 6.2 seconds, corresponding to an average speed between 1.1 and 1.4 m/s. In Izmir, Turkey, Onelcin and Alver (2017) observed six signalized intersections, on which pedestrians crossed four-lane divided highways. The crosswalk length ranged between 13.6 and 22.5 m. For individuals, average crossing speeds between 1.23 and 1.44 m/s were reported, with an overall standard deviation of 0.26 m/s. A walking speed beyond 2.0 m/s was rarely observed. Observational findings regarding crossing speeds are summarized in Figure 5.

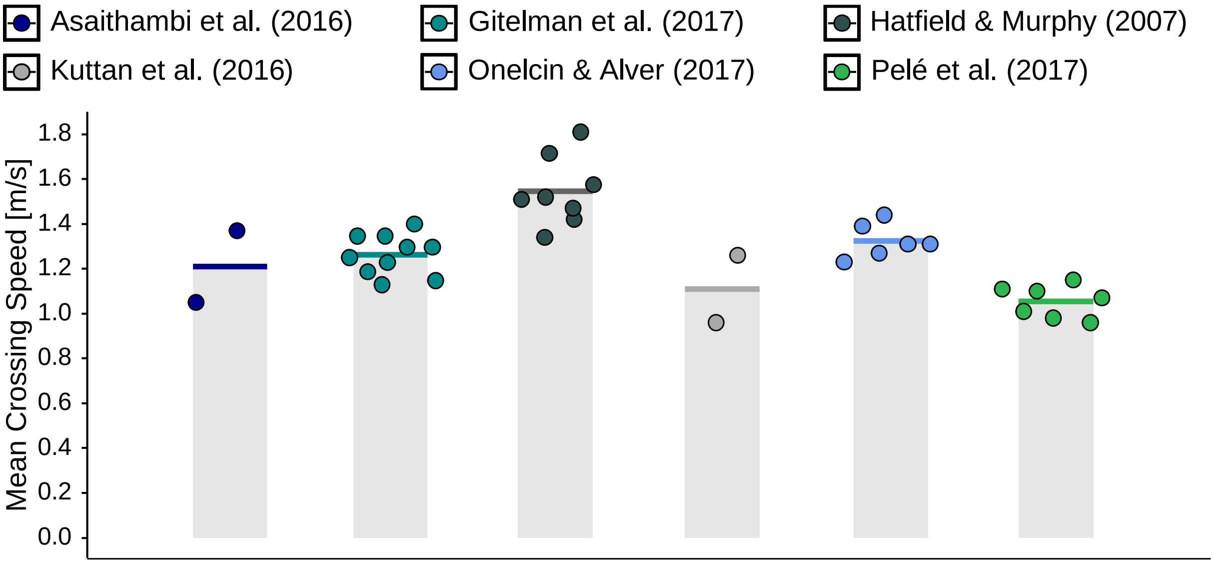

Figure 5 – Average crossing speed in naturalistic traffic according to different publications. Bars represent the median value for the respective publication.

5.3 Further Measures

In contrast to walking speed and gap acceptance, CITs and PETs are analyzed less frequently. Onelcin and Alver (2017) report PETs for individuals to average at 7.82 s, with a standard deviation of 5.26 s. The smallest observed value corresponded to 1.19 s and safety margins tended to be higher for females, seniors, and pedestrians who carried items (Onelcin & Alver, 2017). Focusing on conflict avoidance in a large intersection in Shanghai, China, Tageldin et al. (2017) report average PETs between 0.8 and 1.2 s. These relatively small values, however, must be considered in the context of the variable movements and the large number of traffic participants, resulting in a scenario whose complexity clearly surpasses that of the experimental research presented in this thesis.

Estimates for CIT, lastly, are provided by Oxley et al. (1997), who analyzed crossing behavior in different age groups in Melbourne, Australia. Young pedestrians demonstrated relatively similar CITs of -0.05 and -0.11 s in uni- and bidirectional traffic, respectively. Crossing in the elderly, in contrast, was delayed in bidirectional traffic, resulting in CITs of 0.87 rather than -0.10 s.

6 Summary of Publications

The five publications that form the present thesis can be divided into three parts. First, a procedure for establishing behavioral validity in traffic simulators is formalized, from which recommendations for future research are deduced subsequently (Schneider & Bengler, 2020a). The goal is to thereby devise guidelines, which are applicable to various types of simulators and oriented towards a comparison to uninstructed behavior in naturalistic traffic. The second publication outlines the current state of VR pedestrian research (Schneider & Bengler, 2020b). In this article, a scheme is proposed for the classification of research areas, experimental tasks, and technological setups, which serves as a foundation for identifying relevant questions and scenarios in the subsequent experiments. Together, these two articles form a methodological framework, in which I identify and integrate relevant validation techniques, experimental tasks, and research questions.

The second part of this thesis deals with the replacement of physical test beds by VR technology. Building on the classification derived in the first part, a high-end CAVE-like setup is considered in addition to a non-virtual test track environment to account for different simulator types. Hence, I investigate differences that emerge from the use of VR technology in general, as well as the consequences of employing one of two hardware setups that are common in recent pedestrian research. The corresponding experiment, in which an identical street crossing task is implemented in all three environments, is presented in the third article (Schneider et al., 2021).

To address the apparent artificiality of laboratory settings, the final part of this thesis concerns scenario-dependent differences between naturalistic traffic and VR experiments. Building on a model of human information processing, two simulator studies are conducted to assess the effects of distraction, motivational state, and scenario complexity. These two experiments are outlined in the fourth (Schneider & Li, 2020) and fifth article (Schneider et al., 2019), respectively. The fourth article furthermore includes a discussion of the theoretical foundation for expecting a mismatch between naturalistic and experimental settings and suggestions to support the design of virtual scenarios in which cognitive demands better approach the requirements of real-world traffic.

The resulting outline is summarized in Figure 6. Allocating the three parts from left to right reflects their order of appearance in the scope of this thesis and is not supposed to suggest that technological factors should be perceived as prior to the evaluation of contextual features.

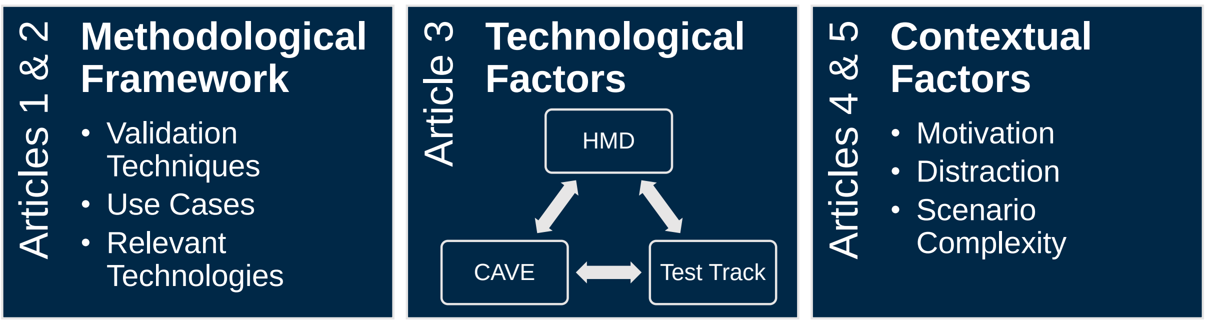

Figure 6 – Outline of the present work. This thesis addresses three objectives, which concern the development of a methodological framework as well as the empirical evaluation of the impact of both technological and contextual factors.

Evaluating Behavioral Validity in Traffic Simulators

Schneider, S., & Bengler, K. (2020). Evaluating behavioral validity in traffic simulators. In T. Lusikka (Ed.), *Proceedings of TRA2020, the 8th Transport Research Arena: Rethinking transport – towards clean and inclusive mobility*, p. 130. (Conference cancelled)

The objective of the present work was to validate VR as a methodology for pedestrian research. Hence, a first step consisted in summarizing and synthesizing existent validation approaches. The resulting overview was published in a first article along with recommendations on best practices. In order to incorporate all relevant techniques and broaden the range of applications, these considerations are not limited to the domain of pedestrian simulators, but equally concern other types of road users.

The first part of this article presents an overview of previous validation approaches and a discussion of their respective benefits and limitations. In line with the observations of Wynne et al. (2019), presumable indicators of validity were found to encompass a broad spectrum of different parameters, ranging from subjective reports of personality traits (Schwebel et al., 2008) to recordings of uninstructed real-world behavior (Jamson, 2001). Approaches identified during the literature survey are listed in Figure 7.

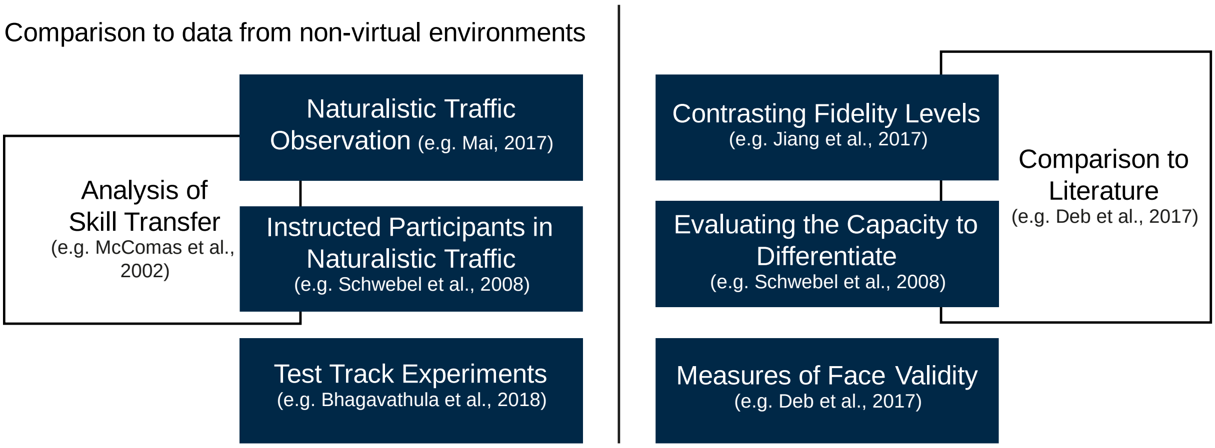

Figure 7 – Overview of methodological approaches for simulator validation. Approaches listed on the left depend on comparison data from non-virtual environments, whereas those on the right do not. The overlap of rectangles indicates techniques that are frequently combined. Skill transfer, for instance, is often assessed by instructing participants to perform the actions of interest in naturalistic traffic after training.

Best practices for future validation studies encompass a deliberate definition of context variables, statistical methods, and thresholds depending on the simulator's purpose. In particular, I discuss the common yet incorrect assumption that the absence of statistical significance provides evidence for the agreement between a simulator and a non-virtual environment (Bellem et al., 2016; Mullen et al., 2011, 13:3-13:4) and outline adequate alternatives for statistical modelling. Three categories are distinguished to describe the purpose of a simulator. The latter include training and education, the contrasting of user characteristics, infrastructure, or technology, and the prediction of behavior in terms of absolute values. Platforms to collect real-world comparison data are suggested based on the affiliation with one of these categories, the riskiness of the scenarios to be investigated, and the observability of dependent measures. The article concludes with two contrasting examples, illustrating the application of the proposed framework.

Importantly, the techniques suggested for simulator validation target the predictive value for uninstructed behavior in real-world traffic, rather than the equivalence of experimental data from virtual and non-virtual laboratories. To account for the unequal accuracy required by different applications, the common dyad of physical and behavioral validity is extended by the concept of simulator appropriateness. Instead of focusing on the agreement of simulated and real environments, the latter refers to the usefulness of a simulator for investigating a particular research question. This additional concept is meant to highlight that although “validity relates to the suitability of the simulation for its intended application” (Allen et al., 2011, 2:9), behavioral agreement and suitability are not synonymous. The relationship between behavioral validity and simulator appropriateness can be described in a similar way as the relationship between physical and behavioral validity: While a meaningful answer to any particular question requires a certain extent of agreement with real-world behavior, the respective minimum varies between applications. Two modalities for alerting a driver, for instance, may reasonably be compared in a simulator that evokes unrealistic braking due to the inertia of the pedal. For the prediction of safety-critical parameters, in contrast, such as whether or not the distance that remains to the preceding car will be sufficient, the same simulator appears inappropriate.

In addition to completing the literature review and writing the manuscript, my personal contribution to this article consisted in summarizing and synthesizing the different approaches and highlighting their specific strengths and weaknesses. Building on this analysis, I deduced recommendations for the selection of validation techniques.

Virtually the Same? Analysing Pedestrian Behaviour by Means of Virtual Reality

Schneider, S. & Bengler, K. (2020): Virtually the same? Analysing pedestrian behaviour by means of virtual reality. *Transportation Research Part F: Traffic Psychology and Behaviour 68*, pp. 231-256.

The second article pertains to the use of VR applications for investigating pedestrian behavior. Based on a literature review of 87 studies published since 2008, I identify categories to classify research with regard to the scientific objective, the technological setup, and the experimental task (Figure 8). Importantly, this classification is derived empirically rather than theoretically. The resulting overview is intended to facilitate the synthesis of existent findings, but also to highlight current gaps from which future research can develop.

The classification of research objectives yielded three high-level categories. In addition to the comparison of street environments and user groups, a relatively large proportion of recent studies was dedicated to the evaluation of VR technology for the purpose of pedestrian research and training. While this finding highlights the interest in establishing VR as a research methodology, approaches vary broadly, and thorough validation studies remain scarce. To the extent that systematic comparisons between virtual and non-virtual environments have been reported, they seem to confirm the pattern of sufficient relative but limited absolute validity, which was previously observed in driving simulation (Blana, 1996, pp. 40–41; Mullen et al., 2011, 13:14). In addition, a similarly ambiguous relationship emerged between physical fidelity and behavioral validity.

The relative frequency of investigations related to the street environment increased towards the end of the considered time period. Such a trend may result from advances in vehicle automation, which increasingly shift the focus of automobile engineering towards the interaction with vulnerable road users (Millard-Ball, 2018). This assumption is supported by a number of related studies, which were published after the review had been completed (de Clercq et al., 2019; Deb et al., 2019; Dietrich et al., 2020; Fuest, Maier et al., 2020; Nuñez Velasco et al., 2019). Until now, however, such investigations mainly target young to middle-aged, healthy individuals, mostly neglecting the diversity of pedestrians.

Experimental tasks were subdivided according to different stages of mental processing. Across all stages, instructions related to street crossing were predominant, with two thirds of all studies including either actual crossing movements or an indication of crossing intent. Since these tasks require little time and most researchers aim to maximize experimental control, current research is dominated by relatively simple and artificial tasks, such as repeatedly crossing between a line of unidirectional traffic. Such experiments may foster an unnatural focus of attention. Accordingly, the current design of scenarios may fail to account for the complexity of real-world traffic, compromising the behavioral validity of pedestrian research.

The introduction of commercial HMDs such as the Oculus Rift and HTC Vive headsets may be the reason for a general rise in publications between 2014 and 2015. In line with that, the proportion of measures that are directly observable by the experimenter (in contrast to those which target internal processes such as perception and decision-making) increased over the past years. Such a development can be explained by the enhanced capacities for naturalistic walking in modern simulators. However, since the physical space available for movements is still restricted, the option for actual walking may indeed limit the range of applicable scenarios, further narrowing their bandwidth in favor of repetitive street crossings.

Figure 8 – Categories to classify VR-based pedestrian research, which were derived in the scope of the second article.

In addition to providing a general framework for the VR-based investigation of pedestrian behavior, the second article formed the foundation for identifying research gaps and selecting use cases and dependent variables in the subsequent experiments. My personal contribution included the selection of publications, the development of the classification scheme, and writing the manuscript.

Pedestrian Crossing Decisions in Virtual Environments: Behavioral Validity in CAVEs and Head-Mounted Displays

Schneider, S., Maruhn, P., Dang, N.-T., Pala, P., Cavallo, V., & Bengler, K. (2021). Pedestrian Crossing Decisions in Virtual Environments: Behavioral Validity in CAVEs and Head-Mounted Displays. *Human Factors* (advance online publication).

The third article targets the influence of VR technology on pedestrian behavior in controlled experiments. To this objective, the HMD-based and the CAVE-like simulator described in Chapter 4.2 were contrasted to a physical test track. These setups were considered representative of the two types of highly immersive simulators, which had been identified in the previous literature review (Schneider & Bengler, 2020b). An identical experimental task, requiring a decision to cross between two approaching vehicles, was presented to thirty adults in each of the environments. Participants were instructed to signal their crossing intent by stepping forward.

The comparison of the three environments revealed noteworthy differences. Across all speeds and gap sizes, most gaps were accepted on the test track (47.5%), although an overall acceptance rate of 45.0% in the CAVE closely approached this value. In the HMD-based simulator, in contrast, participants were generally more hesitant, and acceptance rates dropped to 36.0%. Contrasting previous comparisons between simulator types, which reported more gaps to be accepted in HMDs than in CAVEs (Cavallo, Dang et al., 2019; Mallaro et al., 2017), this observation may highlight a particularity of experimental tasks that target crossing intent rather than actual crossings. Considering the inconsistency with the results of Feldstein and Dyszak (2020), it may furthermore suggest that a general reluctance applies to decisions in HMD-based VR, regardless of the content of this decision.

Higher reluctance was also evident in the timing of crossing initiation. While participants on the test track typically stepped forward before the first vehicle had completely passed them, crossing was delayed by approximately 0.4 s in both simulators. In combination with similar acceptance rates as on the test track, this delay led to a relatively high percentage of PET estimates below zero in the CAVE. These negative PETs indicate a hypothetical collision if participants had actually crossed the street (11.5% of all crossings in the CAVE vs. 2.8% in the HMD and 3.8% on the test track). Curiously, CITs increased with gap size in the CAVE, whereas no such relationship was found in the other environments. In VR, participants seemed more susceptible to changes in vehicle speed, which significantly affected acceptance rates and estimated PETs in both simulators, but not on the test track. The fact that participants in VR accepted smaller gaps at higher speed implies an overreliance on spatial rather than temporal intervals (Feldstein & Dyszak, 2020), which may result from impoverished visual information (Cavallo & Laurent, 1988; Feldstein, 2019).

In a post-experimental questionnaire, participants in the HMD group demonstrated awareness for their reluctance by indicating that they behaved more cautiously than usually and took more time to make a decision. Participants on the test track, in contrast, were more confident, judging their decisions as safer and collisions as less probable than in both simulators. Self-reports may thus support the interpretation of observational data. Interestingly, however, differences in the rating of the danger caused by a hypothetical collision turned out insignificant. This finding may result from an attempt to comply with the instruction to mimic an actual crossing situation and highlights the difficulty in the assessment of risk perception in simulator experiments.

Importantly, this study targeted a comparison of alternative experimental settings. While it demonstrates multiple differences between VR environments and a test track, it does not provide information on the extent to which either setting reproduces the affordances of real-world traffic. In particular the simplified scenario may reasonably be expected to affect the measures under investigation, as outlined in the fourth article (Schneider & Li, 2020). In line with that, several participants highlighted a lack of contextual information in all three environments.

Nonetheless, the deviations between the results obtained in virtual environments and on a test track indicate additional restrictions when predicting real-world pedestrian behavior based on studies conducted in VR. Although technological advancements may counteract certain perceptual biases (Feldstein et al., 2020; Kelly et al., 2017), current simulators still seem insufficient to accurately replicate the sensory input provided by non-virtual stimuli, and this lack of fidelity appears to affect common measures of pedestrian behavior. Due to numerous advantages, VR is likely to remain a popular and valuable tool in pedestrian research. However, biases in information processing and observable responses – such as those observed in the present experiment – must be taken into account when interpreting experimental findings.

My personal contribution to this article included the processing and analysis of the experimental data as well as writing the manuscript. In addition, I was involved in data collection and the experimental design.

Virtual Scenarios for Pedestrian Research: A Matter of Complexity?

Schneider, S. & Li, G. (2020). Virtual Scenarios for Pedestrian Research: A Matter of Complexity? In J. Y. C. Chen & G. Fragomeni (Eds.), *Virtual, Augmented and Mixed Reality. Design and Interaction. HCII 2020. Lecture Notes in Computer Science 12190,* pp. 171 – 190. Springer, Cham.

In the fourth article, I focus on likely differences between natural and simulated environments with respect to human information processing. The underlying research question is related to the concept of ecological validity (Newton et al., 2014), which concerns the extent to which simulated scenarios are representative of the range of events occurring in naturalistic traffic. As outlined in Chapter 2.2, the relationship between physical and behavioral validity is a frequent topic of discussion. Non-technological influences, in contrast, tend to be neglected, but can similarly be assumed to alter human behavior in experimental settings. In driving simulation, for instance, mental workload and gaze behavior were found to deviate from on-road observations (Wynne et al., 2019), likely indicating a mismatch in cognitive processing.

In the first part of this publication, I outline how biases in attentional focus and an elevated predictability of events may affect experimental outcomes. In fact, simplistic and repetitive experimental scenarios seem to systematically exclude some of the most important causes for safety-critical events. In particular for pedestrians, a failure to “observe attentively” and “recognize all the relevant information” (Otte et al., 2012, pp. 151–152) accounts for more than 50% of injuries of all severity levels. In this regard, pedestrians differ from cyclists and motorcyclists, for whom errors in the evaluation of perceived information and maneuver planning are more common (Otte et al., 2012). Accident risk in pedestrians can thus largely be attributed to distraction and misguided observation strategies (Otte et al., 2012), which are mostly irrelevant if their only task consists in monitoring a single and highly predictable line of approaching vehicles. Even if unexpected events are purposefully added, Caird and Horrey (2011, 5:7) claim that “you can only truly surprise a research participant once”, and that even this assumption may be overly optimistic in many experimental settings. Biased expectation arising from a limited range of possible events may thus influence outcome measures such as reaction times, visual monitoring, and risk taking.

In a second part of this publication, I discuss possible mechanisms that might alleviate some of the problems outlined before. Techniques can be grouped into the three categories of minimizing repetitiveness, adding distractors, and varying elements of the experimental conditions to reduce predictability. While the application of such measures relies on empirical data demonstrating their effectiveness and each approach may be infeasible in any particular setting, external validity should not be disproportionally traded to maximize internal validity (Persson & Wallin, 2012). This is particularly true under the assumption that internal validity does not depend on artificiality, but on the awareness for potential confounders (Jimenez-Buedo & Miller, 2010). Consequently, overly simplified scenarios may ultimately increase rather than prevent the ambiguity of results by introducing additional confounders such as boredom, fatigue, and learning effects.

The article concludes by a simulator study, which was designed to evaluate how an increase in complexity may affect common measures of pedestrian behavior. One group of twenty participants experienced a conventional street crossing scenario, in which discontinuous lines of vehicles approached at a uniform speed of 30 km/h. Furthermore, they were informed if any of the vehicles would yield. In a second group of equal size, yielding and non-yielding trials alternated randomly and approaching vehicles changed their speed and direction up to a distance of 40 m from the participant's viewpoint. Additionally, virtual pedestrians walking on both sidewalks, vehicles driving on the neighboring roads, and urban background sounds served as distractors. Differences between the two groups were designed to include several techniques that may lower the artificiality of the experimental scenario while maintaining a relatively high degree of experimental control.

Descriptively, perceived complexity appeared lower in the conventional scenario, whereas no difference was evident with regard to the ratings of realism or predictability. Mixed regression analyses did not reveal any changes in crossing behavior, which was assessed by CIT and crossing speed in all trials and by gap acceptance and PET in trials without yielding vehicles. For non-yielding vehicles and small gaps, participants started crossing earlier and walked faster. The distance at which yielding vehicles started to brake, in contrast, had no significant effect.

Although the observations bear little evidence of differences between the two groups, this must not be understood as proof against behavioral adaptations due to rising complexity. First, although some effects were numerically close to zero, others such as a tendency towards smaller PETs and delayed crossing initiation in yielding trials may have been concealed by interpersonal differences resulting from a between-subject design. Increased variance at small gaps within the group that experienced higher complexity suggests that individuals differ in their capacity to cope with elevated cognitive demands. Since the participant sample almost exclusively comprised young adults with an academic background, such effects may be even more obvious if studying a more heterogeneous population.

Moreover, the experiment was specifically designed to maintain a high degree of experimental control. As a result, the trajectories of approaching vehicles were constant for almost five seconds before they reached the participants. Considering the limited resolution of VR displays, the time during which their speed and distance could accurately be perceived in the conventional scenario may have been similar. Since the chosen distractors were relatively constant and did obviously not interfere with the crossing task, participants may have developed effective strategies to ignore them. Based on the obvious association of attention and traffic safety (Otte et al., 2012), it thus seems more appropriate to understand the findings in terms of the participants' ability to cope with a relatively subtle change in cognitive demands.

Although increased complexity can reasonably be related to road safety, the precise relationship between cognitive demands and pedestrian behavior is unclear. Notably, unrealistic simplification is neither inherent nor restricted to research in virtual environments. In contrast, it may be alleviated by the flexibility they offer for designing virtual scenarios, whereas test tracks often "lack appropriate context" due to "the absence of a visual context and [the] unrepresentative ambient soundscape" (Singh et al., 2015, p. 120). Researchers, however, are often reluctant to make use of this flexibility for concern of experimental control. To recognize the full potential of VR, future studies should thus clarify if naturalistic cognitive demands can be accurately replicated in simulator studies.

My personal contribution to this article included the definition of the research question, the processing and analysis of data, writing of the manuscript, and involvement in the experimental design.

Pedestrian Behavior in Virtual Reality: Effects of Gamification and Distraction

Schneider, S., Ratter, M., & Bengler, K. (2019): Pedestrian Behavior in Virtual Reality: Effects of Gamification and Distraction. In *Proceedings of the Road Safety and Simulation Conference 2019*, Iowa City, Iowa, USA.

As outlined previously, misguided attention is a frequent reason for pedestrian injury (Otte et al., 2012). While the fourth publication concerns the complexity of the traffic scenario, the fifth article pertains to the impact of motivational and cognitive states. Instead of expecting distraction to arise from the presence of external elements, a third simulator experiment was devised to more directly target the associated internal mechanisms.

Cognitive load was induced by an auditory 1-back task, which required neither motoric reactions nor the processing of visual stimuli. This task was specifically selected to interfere as little as possible with the sensory information that formed the virtual scenario. Since pre-tests indicated cognitive load to be elevated but moderate, it was assumed comparable to real-world processes such as a captivating train of thought. Because effects of cognitive load were understood as a performance measure and thus mainly determined by individual capacities, carry-over effects were assumed negligible. Hence, the presence of the secondary task was counterbalanced as a within-subject factor.

In real-world traffic, crossings are usually performed to reach a certain destination, which can be associated with varying degrees of time pressure and motivational states. The expectation that changes in motivation may bias the results of simulator studies has been mentioned previously (Blana, 1996, p. 14). The challenge of empirically testing this assumption, however, stems from the difficulty in manipulating motivational states within an experimental context. In the present article, gamification techniques were employed to this objective. To induce time pressure and motivational incentives, one group of participants experienced a crossing scenario that was enriched by various elements such as a timer, golden coins to be collected while crossing, and a scoring system that rewarded safe but rapid movements.

Thirty-six participants were assigned to one of the two groups. In contrast to the previous experiments, participants were not simply instructed to cross the street safely, but it was additionally pointed out that crossings should be performed as fast as possible. Both groups were presented with continuous traffic. The intervals between vehicles increased until a gap was accepted, upon which the gap size was reset to the initial value while participants were turning their back to the street. The selection of larger gaps thus implies that participants waited longer.

According to a mixed regression analysis, the addition of game elements led to the acceptance of smaller gaps. Although participants compensated for the lack of time by increasing their walking speed, PETs were significantly lower. Due to their overall low number, no inferential analyses were performed on the occurrence of virtual collisions. It is, however, noteworthy, that they did not appear more frequent in the gamified group despite the acceptance of considerably smaller gaps. The overall decline of gap sizes mainly resulted from reduced variability, since larger gaps were rarely accepted in the gamified group. Neither cognitive load itself nor its interaction with gamification had a significant effect on the observed crossing behavior.

Ratings on the Play Experience Scale (Pavlas et al., 2012) increased due to game elements, approaching the values obtained for actual games. The secondary task, in contrast, lowered play experience, albeit to a lesser extent. The Presence Questionnaire (UQO Cyberpsychology Lab, 2004) indicated lower presence if a secondary task was performed, whereas an increase due to gamification failed to reach statistical significance in this case.

Focusing on the effects of time pressure, the results support findings of elevated risk-taking (Charron et al., 2012; Morrongiello, Corbett, Switzer et al., 2015), leading to evasive adjustments in walking speed (Morrongiello, Corbett, Milanovic et al., 2015). Additionally, crossing behavior may have been altered by the sense of progress and meaningfulness, which was found to arise from continuous feedback and gamification elements such as leaderboards and badges (Sailer et al., 2017). As a central motivator of human behavior, the feeling of success may explain the rise in subjective play experience. While the restricted variance of gap sizes suggests changes in decision-making, the reduction of PETs corresponded to less than half the decrease in gap sizes. This asymmetry might be understood as evidence of higher motivation, since more accurate timing and higher walking speeds were obviously effective in attenuating the effects on safety margins. Such an interpretation is supported by informal observations and comments, which suggest stronger personal engagement and more positive emotions in the gamified group. In comparison to naturalistic traffic (Das et al., 2005; Yannis et al., 2013), however, the median size of accepted gaps seems unrealistically small, resulting in walking speeds that would rarely occur in on-road traffic. Although the present data thus provide preliminary evidence that gamification is effective in inducing motivational incentives, it appears to impair the generalizability to real-world traffic. Considering the common association between video games and VR technology, the game-likeness of pedestrian simulators should thus be considered a potential confounder in future research.

The missing influence of a secondary task contradicts findings of delayed reactions under distraction (Banducci et al., 2016; Neider et al., 2010). While it implies that participants fully compensated for the moderate cognitive load, it is unclear if this observation can be generalized to more complex scenarios. According to the questionnaire data, both presence and play experience suffered, indicating a change in personal involvement. The strategy of balancing higher demands at the cost of subjective comfort may work only for a limited time (Hockey, 1997). Moreover, considering the young and predominantly academic participant sample, it may generally be less effective for other populations.

The results indicate that their motivational state affects pedestrians' crossing behavior, highlighting the need for further research regarding associated differences between real-world and experimental settings. The mismatch in subjective and objective measures challenges the presumed association between the feeling of presence and observable actions.

My personal contribution to this article included the definition of the research question, the processing and analysis of data, writing of the manuscript, and involvement in the experimental design.

7 Comparison to Observational Data

Although the latter three publications (Schneider et al., 2019; Schneider et al., 2021; Schneider & Li, 2020) provide insights on the factors that influence the behavior in VR pedestrian simulators, they do not allow an immediate conclusion regarding the comparability to real-world traffic. To facilitate an assessment of the agreement with real-world behavior, Figure 9 and Figure 10 contrast the findings of the experimental studies with the observational data reported in Chapter 5. At a first glance, the data presented in the third (Schneider et al., 2021) and fourth article (Schneider & Li, 2020) do not appear to systematically stand out, whereas the gap acceptance behavior clearly differs in the fifth publication (Schneider et al., 2019). In comparison to the overall variability, the values observed within a given experimental study seem relatively homogeneous across groups.

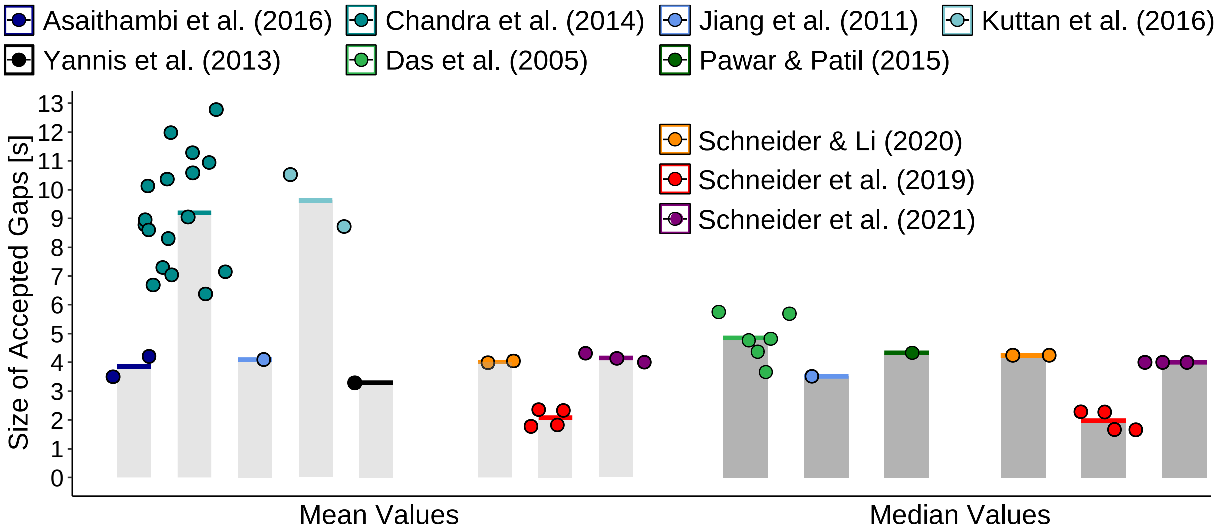

Figure 9 – Comparison of the size of accepted gaps as observed in naturalistic traffic and in the experimental studies conducted in the scope of this thesis. Bars represent the median value for the respective publication. Values on the left refer to the mean accepted gap size, whereas those on the right refer to the median.

The comparison to real-world data once more highlights the impact of scenario design and experimental instructions: In the studies by Schneider et al. (2021) as well as Schneider and Li (2020), the displayed gap sizes were restricted to a maximum of five seconds. Vehicles approached in relatively short and disjoint convoys. While the size of accepted gaps approximately matches real-world observations in those two cases, considerably smaller gaps were accepted according to Schneider et al. (2019). In the latter experiment, gap sizes were designed to slowly increase in continuous traffic and participants were instructed to perform not only safe, but also fast crossings. In line with the acceptance of smaller gaps, the average crossing speed was higher and mean PETs ranged between 0.49 s and 0.72 s, indicating a degree of risk-taking which is unlikely in real world. Notably, although even smaller gaps were accepted in the group who experienced gamification, the observed values clearly differ from the remaining data in both conditions. In addition to the experimental instructions, a reason for this may refer to the effect of waiting time (Asaithambi et al., 2016): Since they had to wait longer before being presented with a 5-s gap, participants in Schneider et al. (2019) may have accepted smaller gaps due to impatience. Hence, although crossing in continuous traffic appears more natural than accepting gaps within disconnected convoys, the observed actions may in fact be less predictive of real-world behavior if gap sizes are not representative.

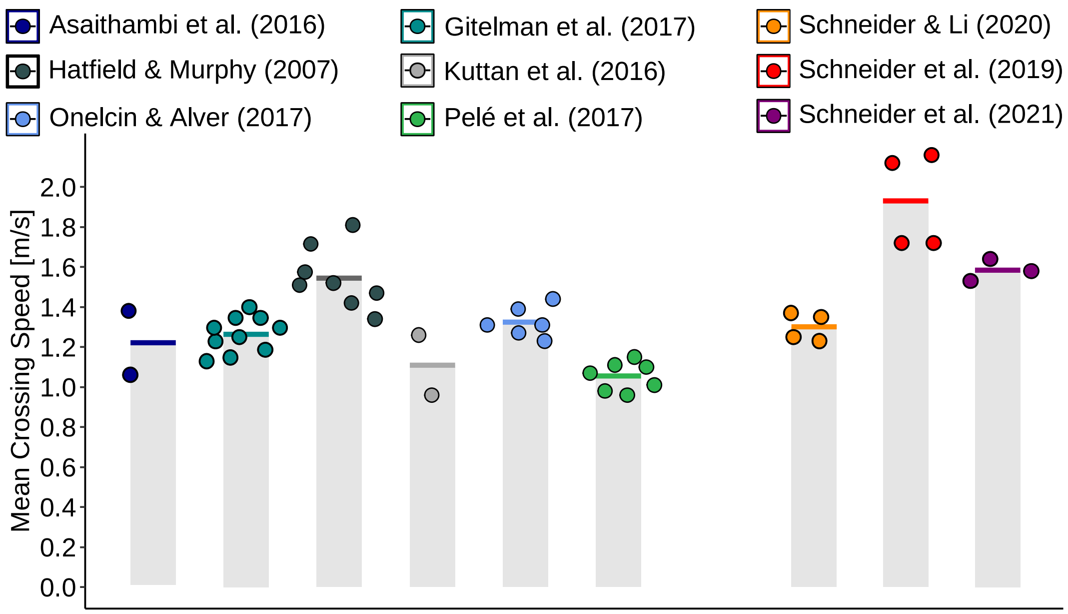

Figure 10 – Comparison of the average crossing speed as observed in naturalistic traffic and in the experimental studies conducted in the scope of this thesis. Bars represent the median value for the respective publication. For Schneider et al. (2021), in which crossing intent was communicated by stepping forward, walking speed was measured prior to the experimental trials. Although these values thus differ from actual assessments of crossing speed, they were included for completeness.

For trials in which vehicles did not yield, CITs in Schneider and Li (2020) averaged at 0.43 s, with no major differences between the groups. While they exceeded the values reported by Oxley et al. (1997) by approximately half a second, this difference may partially be explained by the fact that the threshold for crossing initiation was located 0.25 m from the curb to avoid misclassifications. On the test track (Schneider et al., 2021), a mean CIT of -0.25 s indicates that participants started crossing slightly earlier than in Oxley et al. (1997), which may be explained by a relatively large lateral distance to the vehicle. Regarding the simulators, the mean CIT was 0.25 s in the HMD and 0.41 s in the CAVE, approximately matching the delay reported by Schneider and Li (2020).

PETs in Schneider and Li (2020) averaged at 1.83 s and 2.00 s in the complex and the conventional scenarios, respectively. In Schneider et al. (2021), mean PET estimates corresponded to 1.76 s, 1.63 s, and 1.13 s for the test track, HMD, and CAVE. All these values suggest that safety margins decreased considerably in comparison to the findings of Onelcin and Alver (2017), but tended to be higher than those reported by Tageldin et al. (2017).

8 Discussion with Respect to the State of the Art

8.1 Summary of the Results

Pedestrians received a lot of attention during recent years, and new methodologies have been developed to analyze and predict their actions. While driving simulators have long taken a central place in experimental research (Carsten & Jamson, 2011, p. 87), however, the use of VR to study pedestrians has intensified mostly during the past decade. In order to reach valid conclusions, researchers must know to which extent the data obtained from VR simulators generalize to real-world behavior.

The present thesis contains a framework for classifying the usage of VR pedestrian simulators (Schneider & Bengler, 2020b) along with some key considerations regarding the assessment of behavioral validity (Schneider & Bengler, 2020a). Experimental studies were conducted to contrast common technological configurations (Schneider et al., 2021) and various aspects of scenario design (Schneider et al., 2019; Schneider & Li, 2020). The objective thereby was not only to clarify the extent to which different parameters influence the quality of experimental data, but also to encourage a more comprehensive understanding of the concepts of validity and validation by highlighting the focus on real-world behavior and discussing potential confounders beyond the effects of computer technology.

The empirical findings suggest that both technological and contextual factors affect the experimental outcomes. In comparison to a test track, a higher susceptibility to speed effects in VR and an overall reduced gap acceptance in an HMD-based simulator were observed (Schneider et al., 2021). Regarding scenario design, motivational incentives appeared more effective in changing observable actions than internal or external distractors, although the latter seemed to alter the subjective experience of participants (Schneider et al., 2019; Schneider & Li, 2020). Some contradictions to earlier research can be seen as further evidence for the relevance of scenario design: A decline in gap acceptance in HMDs, for instance, contrasts previous comparisons to both CAVE-like simulators (Mallaro et al., 2017) and physical streets (Feldstein & Dyszak, 2020). Such differences may be attributed to the experimental task, supporting a distinction between actual crossings and the indication of crossing intent (Lobjois & Cavallo, 2009).

The contrasting of experimental findings and existing observational data is complicated by differences in infrastructure and demographics such as culture, age, and gender. Hence, the summary in Chapter 7 is not intended to replace further research facilitating a more direct comparison. Nonetheless, in addition to the observed effects of certain experimental manipulations, differences in comparison to real-world observations should caution researchers to overgeneralize the results obtained in VR. This is particularly true because previous attempts to demonstrate the behavioral validity of pedestrian simulators are limited both in number and in scope (Chapter 2.2), and observations may be biased not only due to impoverished or distorted sensory feedback, but also as a result of an unnatural attentional focus or motivational state.

In particular the evidence for perceptual biases (Schneider et al., 2021) supports the expectation that current simulators are more valuable in deducing relative rather than absolute statements (Feldstein & Dyszak, 2020; Singh et al., 2015). In VR pedestrian simulation, humans apparently find it harder to adequately account for speed variations and delay actions such as crossing initiation. The present work, however, also highlights that such biases appear relatively stable for a given experimental configuration. Hence, they may be accounted for in the interpretation of future data, for instance by assuming a delay in crossing initiation of approximately half a second in comparison to non-virtual environments. Since crossing behavior in VR experiments appears to crucially depend on scenario design, including waiting times and the distribution of available gaps, the proposed mechanisms to approach the affordances of naturalistic traffic may further advance pedestrian research.

8.2 Open Questions and Future Research

Several questions are commonly related to simulator validation but were not addressed in the scope of this thesis. While existing configurations were contrasted in the second and third article, no attempt was made to investigate how the physical fidelity of specific parameters affects behavioral validity. Some research exists on this matter (Chapter 2.2), and the fact that different parameters can be expected to interact with each other (Zöller, 2015) presumably renders their isolated investigation less meaningful. In addition, the low cost of HMDs fosters rapid technological advancements such as increases in resolution and FOV, which will likely be implemented regardless of their relevance for the field of pedestrian research. Although analyzing the effect of specific parameters may still be useful to the appropriate allocation of financial resources, it was therefore no focus of the present work.

In addition to providing the sensory cues which are necessary for perception and action, physical fidelity may be expected to support the feeling of presence, i.e. the impression that virtual elements are real (Deniaud et al., 2015). Similar expectations are related to the concept of face validity, which represents the subjective verisimilitude of a simulation (Schwebel et al., 2008). Although face validity is mostly assumed subordinate to the agreement of objective measures, it may affect the experimental results by altering a subject's motivation (Blana, 1996, p. 8; de Winter et al., 2012). Until now, however, there is little evidence to support the claim that presence or perceived realism actually enhance the agreement of observable actions (Deniaud et al., 2015). In the present thesis, descriptively higher presence ratings were elicited by gamification (Schneider et al., 2019), negating a positive relationship between common questionnaires and the predictive value for real-world behavior.

Nonetheless, the implicit assumption persists that by maximizing physical fidelity, one may "trick" participants into believing that what they experience is real and into reacting accordingly. Carsten and Jamson (2011, p. 93), for instance, argue that "[s]imulator driving is by definition an attempt at convincing participants that they are engaged in an analogue of real-world driving" and "[t]he success with which that is achieved will determine the validity of a given simulator". Even if a simulation perfectly mimicked all sensory information of a real-world context, however, most participants would be unlikely to forget about the experimental setting. Motivational factors, such as an intention to comply with the experimental instructions and the lack of time pressure (Carsten & Jamson, 2011, p. 89), but also the consequences of altered risk perception (Caird & Horrey, 2011, 5:7), are likely to influence their behavior regardless of physical fidelity. By acknowledging the awareness for the experimental context and exploring its implications for the measures of interest, researchers may develop techniques to counteract these biases or account for them in the interpretation of experimental data.

For a number of reasons, no naturalistic observations were conducted in the scope of this thesis. First, obtaining precise information on many relevant measures would have been infeasible. For instance, privacy regulations that apply to video recordings in public spaces in Germany would have prevented the reliable extraction of information such as the approximate age of a crossing pedestrian. The latter, however, was shown to influence common measures such as waiting times, crossing speed (Asaithambi et al., 2016), and CIT (Oxley et al., 1997).

If human observers had manually collected such information, it would have severely reduced the number of observations, leading to further confounders due to a lack of matching data points. Information on distraction and motivational state, which were the focus of the fourth and fifth article, would additionally have depended on survey methods, which can be subject to observer bias and a lack of standardization.

In line with previous research, virtual scenarios contained either dense traffic on a single lane (Schneider et al., 2019; Schneider & Li, 2020), or a very limited number of vehicles (Schneider et al., 2021). Both configurations are unlikely to occur in real world, questioning the availability of adequate comparison data. Finally, as outlined in Chapter 3, the intention was to identify factors that bias experimental observations. Since technological advancements continuously increase both physical fidelity and the flexibility in scenario design, the effort to collect a sufficiently large dataset to contrast a specific simulator configuration seemed disproportionate. Consequently, with regard to the congruence to naturalistic traffic, it seemed more appropriate to restrict the analysis to existing data.

In general, the comparison to real-world traffic is complicated, since the measures obtained from naturalistic observations do not necessarily reflect the experimental variables of interest. Simulator studies frequently target perceptual and cognitive processes, such as anticipating the movement of an approaching vehicle (Feldstein & Dyszak, 2020; Fuest, Maier et al., 2020). Such internal processes cannot be extracted from video observations, but must be inferred from actions like crossing initiation. Walking speed, in contrast, which is commonly analyzed in naturalistic traffic (Asaithambi et al., 2016; Hatfield & Murphy, 2007; Kuttan et al., 2016), is less relevant to the research questions examined in simulators. While a comparison between observational and experimental data can thus help to unveil obvious differences, the actual value of simulator studies may lie in the investigation of those variables that cannot be obtained from videos.

While possible effects of scenario complexity were discussed in the fourth publication (Schneider & Li, 2020), future research should clarify the impact of the experimental design. For instance, there is the frequent assumption of a relatively constant critical time gap, below which pedestrians are not expected to cross (Mai, 2017, p. 5). Instructed participants, in contrast, may understand that they are expected to accept at least some gaps regardless of their size. A systematic variation of gap sizes may provide information on the extent to which the overall distribution influences the acceptance of a given gap. More sophisticated algorithms may furthermore mimic the reactions of human drivers, who adjust their trajectory in response to a pedestrian’s actions (Jiang et al., 2011). Reciprocal adjustments are likely to raise the demands on information processing and allow the investigation of more complex interactions. Alternatively, multiple humans may act simultaneously within the same simulation (Jiang et al., 2016). Although Mestre (2017) argues that HMDs are less suitable to serve as a collaborative platform than CAVEs, connected simulators, in which multiple users can perceive each other’s actions, do exist (Lehsing et al., 2016) and are likely to profit from further technological refinements. Technological progress may also facilitate the consideration of additional variables such as gaze behavior (Dong et al., 2020; Zito et al., 2015) thanks to built-in eye-tracking devices in recent HMDs. Alternatives for position tracking and the control of translational and rotational movements may broaden the range of observable measures, including alternatives to perpendicular crossings (Kuttan et al., 2016).

8.3 Conclusion

The present work targets various questions related to the objective of simulator validation. In the analysis of the chances and challenges of using VR for pedestrian research, an attempt was made to reconcile seemingly contradictory objectives, such as the requirements of external and internal validity (Jimenez-Buedo & Miller, 2010; Persson & Wallin, 2012). Although focusing on the usage of recent HMDs, the proposed classification framework (Schneider & Bengler, 2020b), the considerations on the importance of scenario design (Schneider et al., 2019; Schneider & Li, 2020), and the summary and discussion of validation methods (Schneider & Bengler, 2020a) equally extend to other simulator types.

At a first glance, the term "validation" appears to imply a set of techniques that researchers or technicians can apply to an apparatus in order to render it "valid". What it actually refers to, in contrast, is the process of measuring the agreement between selected parameters that were obtained in different settings. Standardized test criteria exist for the evaluation of flight training simulators (Carsten & Jamson, 2011, p. 94) and for computational algorithms mimicking crowd behavior (Zhou et al., 2014). Although the presumable benefits of a standardized test procedure were discussed also with respect to driving simulation (Blana, 1996, p. 7; Carsten & Jamson, 2011, p. 94), however, such tests are still absent. A likely explanation for this is that the heterogeneity of applications prevents them in both driving and pedestrian research – even though the widespread focus on crossing scenarios may facilitate such standardization to some extent.

VR is an especially versatile tool and its application in pedestrian research is certainly not restricted to the relatively simple street crossing task that was investigated in the present work. Due to their vulnerability in case of collisions, public reports about pedestrians are often limited to crash and injury statistics (National Highway Traffic Safety Administration, 2018). This point of view, however, neglects not only the mechanisms that prevent accidents, but also the importance of comfort and perceived safety to encourage active transportation. In this regard, pedestrian simulators may support our understanding of interpersonal differences and preferences. To exploit their full potential, one must be aware that there is more to validation than the assessment of physical fidelity. While Kaptein et al. (1996, p. 32) argue that "any individual driving simulator needs to be validated", the same may be true for any individual experimental task and scenario. Hence, researchers should consider the parameters that are relevant (albeit not necessarily observable) for pedestrian behavior in naturalistic traffic, and the conditions under which conclusions regarding these parameters may be generalized (Caird & Horrey, 2011, 5:8).

The present work contributes to the ongoing refinement of VR-based traffic research by evaluating various factors that may cause pedestrian behavior in VR to differ from the real world. It thereby provides a foundation for assessing the impact of both technological and contextual aspects in the interpretation of experimental data. While systematic biases appear to exist, their consistency can be seen as promising, since it allows to account for them in the future. In this respect, the findings of this thesis further broaden the opportunities to enhance the safety and comfort of pedestrians by means of VR studies.

References

Agarwal, R. (2019). *Validation of a Pedestrian Simulator for interaction between Pedestrians and Autonomous Vehicles* [Master Thesis]. Delft University of Technology, Delft, Netherlands.

Ahmed, S., Huda, M. N., Rajbhandari, S., Saha, C., Elshaw, M., & Kanarachos, S. (2019). Pedestrian and Cyclist Detection and Intent Estimation for Autonomous Vehicles: A Survey. *Applied Sciences*, *9*(11), 2335. https://doi.org/10.3390/app9112335

Allen, R. W., Rosenthal, T. J., & Cook, M. L. (2011). A Short History of Driving Simulation. In D. L. Fisher, M. Rizzo, J. Caird, & J. D. Lee (Eds.), *Handbook of Driving Simulation for Engineering, Medicine, and Psychology* (2-1-2-11). CRC Press.

Allerton, D. (2009). *Principles of Flight Simulation*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470685662

Asaithambi, G., Kuttan, M. O., & Chandra, S. (2016). Pedestrian Road Crossing Behavior Under Mixed Traffic Conditions: A Comparative Study of an Intersection Before and After Implementing Control Measures. *Transportation in Developing Economies*, *2*(2), 839. https://doi.org/10.1007/s40890-016-0018-5

Atash, F. (1994). Redesigning Suburbia for Walking and Transit: Emerging Concepts. *Journal of Urban Planning and Development*, *120*(1), 48–57. https://doi.org/10.1061/(ASCE)0733-9488(1994)120:1(48)

Banducci, S. E., Ward, N., Gaspar, J. G., Schab, K. R., Crowell, J. A., Kaczmarski, H., & Kramer, A. F. (2016). The Effects of Cell Phone and Text Message Conversations on Simulated Street Crossing. *Human Factors*, *58*(1), 150–162. https://doi.org/10.1177/0018720815609501

Banton, T., Stefanucci, J., Durgin, F., Fass, A., & Proffitt, D. (2005). The Perception of Walking Speed in a Virtual Environment. *Presence: Teleoperators and Virtual Environments*, *14*(4), 394–406. https://doi.org/10.1162/105474605774785262

Bellem, H., Klüver, M., Schrauf, M., Schoner, H.-P., Hecht, H., & Krems, J. F. (2016). Can We Study Autonomous Driving Comfort in Moving-Base Driving Simulators? A Validation Study. *Human Factors*, 442-456. https://doi.org/10.1177/0018720816682647

Bernhard, M., Grosse, K., & Wimmer, M. (2011). Bimodal task-facilitation in a virtual traffic scenario through spatialized sound rendering. *ACM Transactions on Applied Perception*, *8*(4). https://doi.org/10.1145/2043603.2043606

Berton, F., Hoyet, L., Olivier, A.-H., Bruneau, J., Le Meur, O., & Pettré, J. (2020). Eye-Gaze Activity in Crowds: Impact of Virtual Reality and Density. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*.

Bhagavathula, R., Williams, B., Owens, J., & Gibbons, R. (2018). The Reality of Virtual Reality: A Comparison of Pedestrian Behavior in Real and Virtual Environments. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 2056–2060. https://doi.org/10.1177/1541931218621464

Blaauw, G. J. (1982). Driving Experience and Task Demands in Simulator and Instrumented Car: A Validation Study. *Human Factors*, *24*(4), 473–486. https://doi.org/10.1177/001872088202400408

Blana, E. (1996). Driving Simulator Validation Studies: A Literature Review.

Blender [Computer software]. (2017). www.blender.org

Bühler, M. A., & Lamontagne, A. (2018). Circumvention of pedestrians while walking in virtual and physical environments. *IEEE Transactions on Neural Systems and Rehabilitation Engineering.* Advance online publication. https://doi.org/10.1109/TNSRE.2018.2865907

Caird, J. K., & Horrey, W. J. (2011). Twelve Practical and Useful Questions About Driving Simulation. In D. L. Fisher, M. Rizzo, J. Caird, & J. D. Lee (Eds.), *Handbook of Driving Simulation for Engineering, Medicine, and Psychology* (5-1-5-16). CRC Press.

Carsten, O., & Jamson, A. H. (2011). Driving Simulators as Research Tools in Traffic Psychology. In B. E. Porter (Ed.), *Handbook of Traffic Psychology* (pp. 87–96). Elsevier. https://doi.org/10.1016/B978-0-12-381984-0.10007-4

Cavallo, V., Dang, N.-T., Pala, P., Granié, M.-A., Schneider, S., Maruhn, P., & Bengler, K. (2019). Comparison of HMD and CAVE pedestrian simulators. In *Proceedings of the Road Safety and Simulation Conference,* Iowa City, Iowa.

Cavallo, V., Dommes, A., Dang, N.-T., & Vienne, F. (2019). A street-crossing simulator for studying and training pedestrians. *Transportation Research Part F: Traffic Psychology and Behaviour*, *61*, 217–228. https://doi.org/10.1016/j.trf.2017.04.012

Cavallo, V., & Laurent, M. (1988). Visual information and skill level in time-to-collision estimation. *Perception*, *17*(5), 623–632. https://doi.org/10.1068/p170623

Chandra, S., Rastogi, R., & Das, V. R. (2014). Descriptive and parametric analysis of pedestrian gap acceptance in mixed traffic conditions. *KSCE Journal of Civil Engineering*, *18*(1), 284–293. https://doi.org/10.1007/s12205-014-0363-z

Charron, C., Festoc, A., & Guéguen, N. (2012). Do child pedestrians deliberately take risks when they are in a hurry? An experimental study on a simulator. *Transportation Research Part F: Traffic Psychology and Behaviour*, *15*(6), 635–643. https://doi.org/10.1016/j.trf.2012.07.001

Das, S., Manski, C. F., & Manuszak, M. d. (2005). Walk or wait? An empirical analysis of street crossing decisions. *Journal of Applied Econometrics*, *20*(4), 529–548. https://doi.org/10.1002/jae.791

de Clercq, K., Dietrich, A., Núñez Velasco, J. P., de Winter, J., & Happee, R. (2019). External Human-Machine Interfaces on Automated Vehicles: Effects on Pedestrian Crossing Decisions. *Human Factors*, *61*(8), 1353–1370. https://doi.org/10.1177/0018720819836343

de Winter, J. C. F., van Leeuwen, P. M., & Happee, R. (2012). Advantages and Disadvantages of Driving Simulators: A Discussion. *Proceedings of Measuring Behavior*, *8*, 47–50.

Deb, S., Strawderman, L. J., & Carruth, D. W. (2019). Should I cross? Evaluating interface options for autonomous vehicle and pedestrian interaction. In *Proceedings of the Road Safety and Simulation Conference,* Iowa City, Iowa.

Deniaud, C., Honnet, V., Jeanne, B., & Mestre, D. (2015). The concept of “presence” as a measure of ecological validity in driving simulators. *Journal of Interaction Science*, *3*(1), 1–13. https://doi.org/10.1186/s40166-015-0005-z

Dietrich, A., Maruhn, P., Schwarze, L., & Bengler, K. (2020). Implicit Communication of Automated Vehicles in Urban Scenarios: Effects of Pitch and Deceleration on Pedestrian Crossing Behavior. In T. Ahram, W. Karwowski, S. Pickl, & R. Taiar (Eds.), *Advances in Intelligent Systems and Computing. Human Systems Engineering and Design II* (Vol. 1026, pp. 176–181). Springer International Publishing. https://doi.org/10.1007/978-3-030-27928-8_27

Dong, W., Liao, H., Liu, B., Zhan, Z., Liu, H., Meng, L., & Liu, Y. (2020). Comparing pedestrians' gaze behavior in desktop and in real environments. *Cartography and Geographic Information Science*, *47*(5), 432–451. https://doi.org/10.1080/15230406.2020.1762513

Engen, T. (2008). *Use and Validation of Driving Simulators* [Doctoral Thesis]. Norwegian University of Science and Technology, Trondheim.

European Commission. (2018). *Road safety in the European Union: Trends, statistics and main challenges*. Publications Office.

Feldstein, I. T. (2019). Impending Collision Judgment from an Egocentric Perspective in Real and Virtual Environments: A Review. *Perception*, *48*(9), 769–795. https://doi.org/10.1177/0301006619861892

Feldstein, I. T., & Dyszak, G. N. (2020). Road crossing decisions in real and virtual environments: A comparative study on simulator validity. *Accident Analysis and Prevention*, *137*, 105356. https://doi.org/10.1016/j.aap.2019.105356

Feldstein, I. T., & Ellis, S. R. (2020). A Simple Video-Based Technique for Measuring Latency in Virtual Reality or Teleoperation. *IEEE Transactions on Visualization and Computer Graphics*, *PP.* https://doi.org/10.1109/TVCG.2020.2980527

Feldstein, I. T., Kölsch, F. M., & Konrad, R. (2020). Egocentric Distance Perception: A Comparative Study Investigating Differences Between Real and Virtual Environments. *Perception*, *49*(9), 940–967. https://doi.org/10.1177/0301006620951997

Focas, C., & Christidis, P. (2017). Peak Car in Europe? *Transportation Research Procedia*, *25*, 531–550. https://doi.org/10.1016/j.trpro.2017.05.437

Fuest, T., Maier, A. S., Bellem, H., & Bengler, K. (2020). How Should an Automated Vehicle Communicate Its Intention to a Pedestrian? – A Virtual Reality Study. In T. Ahram, W. Karwowski, S. Pickl, & R. Taiar (Eds.), *Advances in Intelligent Systems and Computing. Human Systems Engineering and Design II* (Vol. 1026, pp. 195–201). Springer International Publishing. https://doi.org/10.1007/978-3-030-27928-8_30

Fuest, T., Schmidt, E., & Bengler, K. (2020). Comparison of Methods to Evaluate the Influence of an Automated Vehicle's Driving Behavior on Pedestrians: Wizard of Oz, Virtual Reality, and Video. *Information*, *11*(6), 291. https://doi.org/10.3390/info11060291

Gitelman, V., Carmel, R., Pesahov, F., & Chen, S. (2016). Changes in road-user behaviors following the installation of raised pedestrian crosswalks combined with preceding speed humps, on urban arterials. *Transportation Research Part F: Traffic Psychology and Behaviour.* Advance online publication. https://doi.org/10.1016/j.trf.2016.07.007

Gitelman, V., Carmel, R., Pesahov, F., & Hakkert, S. (2017). An examination of the influence of crosswalk marking removal on pedestrian safety as reflected in road user behaviours. *Transportation Research Part F: Traffic Psychology and Behaviour*, *46*, 342–355. https://doi.org/10.1016/j.trf.2016.03.007

Goodenough, R. (2010). *The geometric field of view and speed perception in a driving simulator* [Master's thesis]. Clemson University. http://tigerprints.clemson.edu/cgi/viewcontent.cgi?article=1978&context=all_theses

Grime, G. (1958). Research in human factors in road transport. *Ergonomics*, *1*(2), 151–162. https://doi.org/10.1080/00140135808964582

Habibovic, A., Tivesten, E., Uchida, N., Bärgman, J., & Ljung Aust, M. (2013). Driver behavior in car-to-pedestrian incidents: An application of the Driving Reliability and Error Analysis Method (DREAM). *Accident; Analysis and Prevention*, *50*, 554–565. https://doi.org/10.1016/j.aap.2012.05.034

Haga, S., Sano, A., Sekine, Y., Sato, H., Yamaguchi, S., & Masuda, K. (2015). Effects of using a Smart Phone on Pedestrians' Attention and Walking. *Procedia Manufacturing*, *3*, 2574–2580. https://doi.org/10.1016/j.promfg.2015.07.564

Hatfield, J., & Murphy, S. (2007). The effects of mobile phone use on pedestrian crossing behaviour at signalised and unsignalised intersections. *Accident Analysis & Prevention*, *39*(1), 197–205. https://doi.org/10.1016/j.aap.2006.07.001

Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*, *45*(1-3), 73–93. https://doi.org/10.1016/S0301-0511(96)05223-4

Honey-Roses, J., Anguelovski, I., Bohigas, J., Chireh, V., Daher, C., Konijnendijk, C., Litt, J., Mawani, V., McCall, M., Orellana, A., Oscilowicz, E., Sánchez, U., Senbel, M., Tan, X., Villagomez, E., Zapata, O., & Nieuwenhuijsen, M. (2020). *The Impact of COVID-19 on Public Space: A Review of the Emerging Questions.* https://doi.org/10.31219/osf.io/rf7xa

Horswill, M. S., & Plooy, A. M. (2008). Reducing contrast makes speeds in a video-based driving simulator harder to discriminate as well as making them appear slower. *Perception*, *37*(8), 1269–1275. https://doi.org/10.1068/p5821

Iryo-Asano, M., Hasegawa, Y., & Dias, C. (2018). Applicability of Virtual Reality Systems for Evaluating Pedestrians' Perception and Behavior. *Transportation Research Procedia*, *34*, 67–74. https://doi.org/10.1016/j.trpro.2018.11.015

Jamson, H. (2001). Image Characteristics and Their Effect on Driving Simulator Validity. In *Proceedings of the First International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design: driving assessment 2001* (pp. 190–195). University of Iowa. https://doi.org/10.17077/drivingassessment.1036

Jiang, X., Wang, W., Mao, Y., Bengler, K., & Bubb, H. (2011). Situational Factors of Influencing Drivers to Give Precedence to Jaywalking Pedestrians at Signalized Crosswalk. *International Journal of Computational Intelligence Systems*, *4*(6), 1407–1414. https://doi.org/10.1080/18756891.2011.9727892

Jiang, Y., O'Neal, E. E., Franzen, L., Yon, J. P., Plumert, J. M., & Kearney, J. K. (2017). The influence of stereoscopic image display on pedestrian road crossing in a large-screen virtual environment. *Proceedings - SAP 2017, ACM Symposium on Applied Perception.* Advance online publication. https://doi.org/10.1145/3119881.3119886

Jiang, Y., Rahimian, P., O'Neal, E. E., Plumert, J. M., Yon, J. P., Kearney, J. K., & Franzen, L. (2016). Acting together: Joint pedestrian road crossing in an immersive virtual environment. *Proceedings - IEEE Virtual Reality*, *2016-July.* https://doi.org/10.1109/VR.2016.7504719

Jimenez-Buedo, M., & Miller, L. M. (2010). Why a trade-off? The relationship between the external and internal validity of experiments. *Theoria. International Journal for Theory, History and Foundations of Science*, *25*(3), 301–321.

Kaptein, N., Theeuwes, J., & van der Horst, R. (1996). Driving Simulator Validity: Some Considerations. *Transportation Research Record: Journal of the Transportation Research Board*, *1550*, 30–36. https://doi.org/10.3141/1550-05

Kelly, J. W., Cherep, L. A., & Siegel, Z. D. (2017). Perceived space in the HTC Vive. *ACM Transactions on Applied Perception*, *15*(1), 1–16. https://doi.org/10.1145/3106155

Kuttan, M. O., Babu, S. V., & Asaithambi, G. (2016). Analysis and Modelling of Pedestrian Road Crossing Pattern on Urban Undivided Roads in Mixed Traffic. In *International Conference on Transportation Planning and Implementation Methodologies for Developing Countries,* Mumbai, India.

Lehsing, C., Benz, T., & Bengler, K. (2016). Insights into Interaction - Effects of Human-Human Interaction in Pedestrian Crossing Situations using a linked Simulator Environment. *IFAC-PapersOnLine*, *49*(19), 138–143. https://doi.org/10.1016/j.ifacol.2016.10.475

Leonard, J. J., & Wierwille, W. W. (1975). Human Performance Validation of Simulators: Theory and Experimental Verification. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *19*(4), 446–456. https://doi.org/10.1177/154193127501900412

Lobjois, R., & Cavallo, V. (2007). Age-related differences in street-crossing decisions: The effects of vehicle speed and time constraints on gap selection in an estimation task. *Accident Analysis and Prevention*, *39*(5), 934–943. https://doi.org/10.1016/j.aap.2006.12.013

Lobjois, R., & Cavallo, V. (2009). The effects of aging on street-crossing behavior: From estimation to actual crossing. *Accident Analysis and Prevention*, *41*(2), 259–267. https://doi.org/10.1016/j.aap.2008.12.001

Loudon, K., Zwarenstein, M., Sullivan, F., Donnan, P., & Treweek, S. (2015). Do pragmatic trials trade-off internal validity for external validity? *Trials*, *16*(S2). https://doi.org/10.1186/1745-6215-16-S2-O81

Mai, K. (2017). *Evaluation of PC-Based Virtual Reality as a Tool to Analyze Pedestrian Behavior at Midblock Crossings* [Master's Thesis]. California Polytechnic State University, San Luis Obispo.

Maillot, P., Dommes, A., Dang, N.-T., & Vienne, F. (2017). Training the elderly in pedestrian safety: Transfer effect between two virtual reality simulation devices. *Accident Analysis and Prevention*, *99*, 161–170. https://doi.org/10.1016/j.aap.2016.11.017

Mallaro, S., Rahimian, P., O'Neal, E. E., Plumert, J. M., & Kearney, J. K. (2017). A comparison of head-mounted displays vs. large-screen displays for an interactive pedestrian simulator. In M. Fjeld, M. Fratarcangeli, D. Sjölie, O. Staadt, & J. Unger (Eds.), *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology - VRST '17* (pp. 1–4). ACM Press. https://doi.org/10.1145/3139131.3139171

Maruhn, P., Dietrich, A., Prasch, L., & Schneider, S. (2020). Analyzing Pedestrian Behavior in Augmented Reality — Proof of Concept. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*.

Maruhn, P., Schneider, S., & Bengler, K. (2019). Measuring egocentric distance perception in virtual reality: Influence of methodologies, locomotion and translation gains. *PLOS ONE*, *14*(10), e0224651. https://doi.org/10.1371/journal.pone.0224651

McComas, J., MacKay, M., & Pivik, J. (2002). Effectiveness of virtual reality for teaching pedestrian safety. *CyberPsychology & Behavior*, *5*(3), 185–190. https://doi.org/10.1089/109493102760147150

Mestre, D. R. (2017). CAVE versus Head-Mounted Displays: Ongoing thoughts. *Electronic Imaging*, *2017*(3), 31–35. https://doi.org/10.2352/ISSN.2470-1173.2017.3.ERVR-094

Millard-Ball, A. (2018). Pedestrians, Autonomous Vehicles, and Cities. *Journal of Planning Education and Research*, *38*(1), 6–12. https://doi.org/10.1177/0739456X16675674

Molino, J. A., Opiela, K. S., Katz, B. J., & Moyer, M. J. (2005). Validate First; Simulate Later: A New Approach Used at the FHWA Highway Driving Simulator. In *Proceedings of the Driving Simulation Conference,* Orlando, USA.

Montuwy, A., Cahour, B., & Dommes, A. (2017). Questioning user experience: A comparison between visual, auditory and haptic guidance messages among older pedestrians. *ACM*

International Conference Proceeding Series, *Part F131371.* https://doi.org/10.1145/3125571.3125572

Morrongiello, B. A., Corbett, M., Milanovic, M., Pyne, S., & Vierich, R. (2015). Innovations in using virtual reality to study how children cross streets in traffic: Evidence for evasive action skills. *Injury Prevention*, *21*(4), 266–270. https://doi.org/10.1136/injuryprev-2014-041357

Morrongiello, B. A., Corbett, M., Switzer, J., & Hall, T. (2015). Using a Virtual Environment to Study Pedestrian Behaviors: How Does Time Pressure Affect Children's and Adults' Street Crossing Behaviors? *Journal of Pediatric Psychology*, *40*(7), 697–703. https://doi.org/10.1093/jpepsy/jsv019

Mullen, N., Charlton, J. L., Devlin, A., & Bédard, M. (2011). Simulator Validity: Behaviors Observed on the Simulator and on the Road. In D. L. Fisher, M. Rizzo, J. Caird, & J. D. Lee (Eds.), *Handbook of Driving Simulation for Engineering, Medicine, and Psychology* (13-1 - 13-17). CRC Press.

National Highway Traffic Safety Administration (Ed.). (2018). *Traffic Safety Facts Data 2016 Pedestrians*. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812493

Neider, M. B., McCarley, J. S., Crowell, J. A., Kaczmarski, H., & Kramer, A. F. (2010). Pedestrians, vehicles, and cell phones. *Accident Analysis & Prevention*, *42*(2), 589–594. https://doi.org/10.1016/j.aap.2009.10.004

Newton, P. E., Robinson, N., Lagrange, M., Shaw, S. D., & Assessment, C. (2014). *Validity in Educational and Psychological Assessment*. SAGE Publications Ltd.

Nuñez Velasco, J. P., Lee, Y. M., Uttley, J., Solernou, A., Farah, H., van Arem, B., Hagenzieker, M., & Merat, N. (2019). Interactions with Automated Vehicles: The Effect of Drivers' Attentiveness and Presence on Pedestrians' Road Crossing Behavior. In *Proceedings of the Road Safety and Simulation Conference,* Iowa City, Iowa.

Ogilvie, D., Foster, C. E., Rothnie, H., Cavill, N., Hamilton, V., Fitzsimons, C. F., & Mutrie, N. (2007). Interventions to promote walking: Systematic review. *BMJ (Clinical Research Ed.)*, *334*(7605), 1204. https://doi.org/10.1136/bmj.39198.722720.BE

Onelcin, P., & Alver, Y. (2017). The crossing speed and safety margin of pedestrians at signalized intersections. *Transportation Research Procedia*, *22*, 3–12. https://doi.org/10.1016/j.trpro.2017.03.002

Otte, D., Jänsch, M., & Haasper, C. (2012). Injury protection and accident causation parameters for vulnerable road users based on German In-Depth Accident Study GIDAS. *Accident; Analysis and Prevention*, *44*(1), 149–153. https://doi.org/10.1016/j.aap.2010.12.006

Oxley, J., Fildes, B. N., Ihsen, E., Charlton, J. L., & Day, R. H. (1997). Differences in traffic judgements between young and old adult pedestrians. *Accident Analysis & Prevention*, *29*(6), 839–847. https://doi.org/10.1016/S0001-4575(97)00053-5

Pala, P., Cavallo, V., Dang, N. T., Granié, M.-A., Schneider, S., Maruhn, P., & Bengler, K. (2021). Analysis of Street-Crossing Behavior: Comparing a CAVE Simulator and a Head-Mounted Display among Younger and Older Adults. *Accident Analysis & Prevention*, *152.* https://doi.org/10.1016/j.aap.2021.106004

Papadimitriou, E., Yannis, G., & Golias, J. (2009). A critical assessment of pedestrian behaviour models. *Transportation Research Part F: Traffic Psychology and Behaviour*, *12*(3), 242–255. https://doi.org/10.1016/j.trf.2008.12.004

Pavlas, D., Jentsch, F., Salas, E., Fiore, S. M., & Sims, V. (2012). The Play Experience Scale: Development and validation of a measure of play. *Human Factors*, *54*(2), 214–225. https://doi.org/10.1177/0018720811434513

Pawar, D. S., & Patil, G. R. (2015). Pedestrian temporal and spatial gap acceptance at mid-block street crossing in developing world. *Journal of Safety Research*, *52*, 39–46. https://doi.org/10.1016/j.jsr.2014.12.006

Pelé, M., Bellut, C., Debergue, E., Gauvin, C., Jeanneret, A., Leclere, T., Nicolas, L., Pontier, F., Zausa, D., & Sueur, C. (2017). Cultural influence of social information use in pedestrian road-crossing behaviours. *Royal Society Open Science*, *4*(2), 160739. https://doi.org/10.1098/rsos.160739

Persson, J., & Wallin, A. (2012). Why internal validity is not prior to external validity. In *Philosophy of Science Assoc. 23rd Biennial Mtg,* San Diego CA.

Pinto, M., Cavallo, V., & Ohlmann, T. (2008). The development of driving simulators: Toward a multisensory solution. *Le Travail Humain*, *71*(1), 62. https://doi.org/10.3917/th.711.0062

Pour-Rouholamin, M., & Zhou, H. (2016). Investigating the risk factors associated with pedestrian injury severity in Illinois. *Journal of Safety Research*, *57*, 9–17. https://doi.org/10.1016/j.jsr.2016.03.004

Quintero, R., Parra, I., Lorenzo, J., Fernandez-Llorca, D., & Sotelo, M. A. (2017). Pedestrian intention recognition by means of a Hidden Markov Model and body language. In *IEEE ITSC 2017: 20th International Conference on Intelligent Transportation Systems: Mielparque Yokohama in Yokohama, Kanagawa, Japan, October 16-19, 2017* (pp. 1–7). IEEE. https://doi.org/10.1109/ITSC.2017.8317766

Renner, R. S., Velichkovsky, B. M., & Helmert, J. R. (2013). The perception of egocentric distances in virtual environments - A review. *ACM Computing Surveys*, *46*(2), 1–40. https://doi.org/10.1145/2543581.2543590

Retting, R. (2019). *Pedestrian Traffic Fatalities by State: 2018 Preliminary Data*. https://www.ghsa.org/resources/Pedestrians19

SafetyNet (Ed.). (2009). *Pedestrians & Cyclists.*

Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior*, *69*, 371–380. https://doi.org/10.1016/j.chb.2016.12.033

Schneider, S., & Bengler, K. (2020a). Evaluating behavioral validity in traffic simulators. In T. Lusikka (Ed.), *Proceedings of TRA2020, the 8th Transport Research Arena: Rethinking transport – towards clean and inclusive mobility* (p. 130). https://www.researchgate.net/publication/339746571

Schneider, S., & Bengler, K. (2020b). Virtually the same? Analysing pedestrian behaviour by means of virtual reality. *Transportation Research Part F: Traffic Psychology and Behaviour*, *68*, 231–256. https://doi.org/10.1016/j.trf.2019.11.005

Schneider, S., & Li, G. (2020). Virtual Scenarios for Pedestrian Research: A Matter of Complexity? In J. Y. C. Chen & G. Fragomeni (Chairs), *HCII 2020*.

Schneider, S., Maruhn, P., & Bengler, K. (2018). Locomotion, Non-Isometric Mapping and Distance Perception in Virtual Reality. In *Proceedings of the 2018 10th International Conference on Computer and Automation Engineering - ICCAE 2018* (pp. 22–26). ACM Press. https://doi.org/10.1145/3192975.3193022

Schneider, S., Maruhn, P., Dang, N.-T., Pala, P., Cavallo, V., & Bengler, K. (2021). Pedestrian Crossing Decisions in Virtual Environments: Behavioral Validity in CAVEs and Head-Mounted Displays. *Human Factors.* Advance online publication. https://doi.org/10.1177/0018720820987446

Schneider, S., Ratter, M., & Bengler, K. (2019). Pedestrian Behavior in Virtual Reality: Effects of Gamification and Distraction. In *Proceedings of the Road Safety and Simulation Conference,* Iowa City, Iowa.

Schneider, S., Salloum, M., Gundel, K., & Boos, A. (Accepted). Estimating Time to Contact in Virtual Reality: Does Contrast Matter? In *Proceedings of the 21st Congress of the International Ergonomics Association (IEA 2021),* Vancouver, Canada.

Schram, A. (2005). Artificiality: The tension between internal and external validity in economic experiments. *Journal of Economic Methodology*, *12*(2), 225–237. https://doi.org/10.1080/13501780500086081

Schwebel, D. C., Gaines, J., & Severson, J. (2008). Validation of virtual reality as a tool to understand and prevent child pedestrian injury. *Accident Analysis & Prevention*, *40*(4), 1394–1400. https://doi.org/10.1016/j.aap.2008.03.005

Schwebel, D. C., Severson, J., & He, Y. (2017). Using smartphone technology to deliver a virtual pedestrian environment: Usability and validation. *Virtual Reality*, *21*(3), 145–152. https://doi.org/10.1007/s10055-016-0304-x

Shirazi, M. S., & Morris, B. (2015). Observing behaviors at intersections: A review of recent studies & developments. In *IV2015: 2015 IEEE Intelligent Vehicles Symposium: June 28-July 1, 2015, COEX, Seoul, Korea* (pp. 1258–1263). IEEE. https://doi.org/10.1109/IVS.2015.7225855

Singh, S., Payne, S. R., Mackrill, J. B., & Jennings, P. A. (2015). Do experiments in the virtual world effectively predict how pedestrians evaluate electric vehicle sounds in the real world? *Transportation Research Part F: Traffic Psychology and Behaviour*, *35*, 119–131. https://doi.org/10.1016/j.trf.2015.10.012

Smeed, R. J. (1949). Some Statistical Aspects of Road Safety Research. *Journal of the Royal Statistical Society. Series a (General)*, *112*(1), 1. https://doi.org/10.2307/2984177

Snowden, R. J., Stimpson, N., & Ruddle, R. A. (1998). Speed perception fogs up as visibility drops. *Nature*, *392*(6675), 450. https://doi.org/10.1038/33049

Stoker, P., Garfinkel-Castro, A., Khayesi, M., Odero, W., Mwangi, M. N., Peden, M., & Ewing, R. (2015). Pedestrian Safety and the Built Environment. *Journal of Planning Literature*, *30*(4), 377–392. https://doi.org/10.1177/0885412215595438

Sueur, C., Class, B., Hamm, C., Meyer, X., & Pelé, M. (2013). Different risk thresholds in pedestrian road crossing behaviour: A comparison of French and Japanese approaches. *Accident; Analysis and Prevention*, *58*, 59–63. https://doi.org/10.1016/j.aap.2013.04.027

Tageldin, A., Zaki, M. H., & Sayed, T. (2017). Examining pedestrian evasive actions as a potential indicator for traffic conflicts. *IET Intelligent Transport Systems*, *11*(5), 282–289. https://doi.org/10.1049/iet-its.2016.0066

Tight, M. R., Carsten, O. M. J., & Sherborne, D. (1989). *Problems for vulnerable road users in Great Britain.* Institute of Transport Studies, University of Leeds. http://eprints.whiterose.ac.uk/2267/

Unity [Computer software]. (2018).

UQO Cyberpsychology Lab. (2004). *Presence Questionnaire*. http://w3.uqo.ca/cyberpsy/en/index_en.htm van Houten, R. (2011). Pedestrians. In B. E. Porter (Ed.), *Handbook of Traffic Psychology* (pp. 353–365). Elsevier. https://doi.org/10.1016/B978-0-12-381984-0.10025-6

Wynne, R. A., Beanland, V., & Salmon, P. M. (2019). Systematic review of driving simulator validation studies. *Safety Science*, *117*, 138–151. https://doi.org/10.1016/j.ssci.2019.04.004

Yannis, G., Papadimitriou, E., & Theofilatos, A. (2013). Pedestrian gap acceptance for mid-block street crossing. *Transportation Planning and Technology*, *36*(5), 450–462. https://doi.org/10.1080/03081060.2013.818274

Young, M. K., Gaylor, G. B., Andrus, S. M., & Bodenheimer, B. (2014). A comparison of two cost-differentiated virtual reality systems for perception and action tasks. *Proceedings of the ACM Symposium on Applied Perception, SAP 2014.* Advance online publication. https://doi.org/10.1145/2628257.2628261

Zhou, Y., Klein, W., & Mayer, G. (2014). Guideline for Crowd Evacuation Simulation Validation of a Pedestrian Simulator with RiMEA Test Scenarios. In R. Jahanbegloo (Ed.), *Introduction to nonviolence* (pp. 1–10). Palgrave Macmillan. https://doi.org/10.1007/978-1-137-31426-0_1

Zito, G. A., Cazzoli, D., Scheffler, L., Jager, M., Muri, R. M., Mosimann, U. P., Nyffeler, T., Mast, F. W., & Nef, T. (2015). Street crossing behavior in younger and older pedestrians: An eye- and head-tracking study. *BMC Geriatrics*, *15*, 176. https://doi.org/10.1186/s12877-015-0175-0

Zöller, I. (2015). *Analyse des Einflusses ausgewählter Gestaltungsparameter einer Fahrsimulation auf die Fahrerverhaltensvalidität* [Dissertation]. TU Darmstadt, Darmstadt. http://tuprints.ulb.tu-darmstadt.de/4608/1/20150625_Dissertation_Zoeller_online_PDFX3.pdf

Appendix A – Publications

Schneider, S., & Bengler, K. (2020). Evaluating behavioral validity in traffic simulators. In T. Lusikka (Ed.), *Proceedings of TRA2020, the 8th Transport Research Arena: Rethinking transport – towards clean and inclusive mobility*, p. 130. (Conference cancelled)

Available from https://www.researchgate.net/publication/339746571

Schneider, S., & Bengler, K. (2020). Virtually the same? Analysing pedestrian behaviour by means of virtual reality. *Transportation Research Part F: Traffic Psychology and Behaviour, 68*, pp. 231 – 256.

Available from https://doi.org/10.1016/j.trf.2019.11.005

Schneider, S., Maruhn, P., Dang, N.-T., Pala, P., Cavallo, V., & Bengler, K. (2021). Pedestrian Crossing Decisions in Virtual Environments: Behavioral Validity in CAVEs and Head-Mounted Displays. *Human Factors* (advance online publication).

Available from https://doi.org/10.1177/0018720820987446

Schneider, S. & Li, G. (2020). Virtual Scenarios for Pedestrian Research: A Matter of Complexity? In J. Y. C. Chen & G. Fragomeni (Eds.), *Virtual, Augmented and Mixed Reality. Design and Interaction. HCII 2020. Lecture Notes in Computer Science 12190,* pp. 171 – 190. Springer, Cham.

Available from https://doi.org/10.1007/978-3-030-49695-1_12

Schneider, S., Ratter, M., & Bengler, K. (2019). Pedestrian Behavior in Virtual Reality: Effects of Gamification and Distraction. In *Proceedings of the Road Safety and Simulation Conference*, Iowa City, Iowa, USA.

Available from https://mediatum.ub.tum.de/1537102

Appendix B – Related Work

Besides the five publications that form the present thesis, additional data were collected to evaluate the influence of technological aspects in more detail. In particular, this research targets the effects of VR-related perceptual biases and their interaction with participant characteristics. Since the underlying questions are related to the future evaluation of VR pedestrian simulators, central findings are summarized in the following.

B.1 Distance Compression

Schneider, S., Maruhn, P., & Bengler, K. (2018). Locomotion, Non-Isometric Mapping and Distance Perception in Virtual Reality. In *Proceedings of the 2018 10th International Conference on Computer and Automation Engineering - ICCAE 2018*, Brisbane, Australia, pp. 22 – 26. ACM Press. Available from https://doi.org/10.1145/3192975.3193022

Maruhn, P., Schneider, S., & Bengler, K. (2019). Measuring egocentric distance perception in virtual reality: Influence of methodologies, locomotion and translation gains. *PLOS ONE,* 14(10), e0224651. Available from https://doi.org/10.1371/journal.pone.0224651

The underestimation of distances is a persistent concern in virtual environments (Renner et al., 2013). In pedestrian simulation, it may cause a misjudgment of both the time available for crossing and the street width to cover, and thus lead to suboptimal crossing decisions. Although a number of studies have been performed on this topic, technological advancements and the specific affordances of pedestrian research warrant a reconsideration of previous findings. Hence, two simulator studies (Maruhn et al., 2019; Schneider et al., 2018) were devised to assess the extent of distance compression in VR pedestrian simulators, and to explore potential countermeasures in form of active locomotion.

In a first experiment (Schneider et al., 2018), participants provided verbal estimates of distances between 3 and 4 m in a virtual street environment, which was displayed via an HTC Vive HMD. In one of two experimental blocks, they walked to the target before announcing their decision. These episodes of walking were subject to adjustments in translation gain, varying the ratio between the walked and the visually perceived distance between 90% and 130%. Overall, the observed underestimation of distances was comparable to previous research (Renner et al., 2013). Higher and thus more accurate estimates were obtained after walking to the target, indicating a positive effect of active locomotion. Interestingly, improvements were particularly pronounced if walking was performed in the second experimental block, whereas previous walking experience did not seem to propagate to subsequent static judgments. If the walked distance increased due to a lower translation gain, estimates were higher, highlighting the influence of proprioceptive and vestibular information.

Since verbal estimates may be biased by a misrepresentation of physical units, a second study (Maruhn et al., 2019) was intended to contrast a range of different methods for quantifying distance perception. In line with the first experiment, the objective was to clarify the relative importance of depth cues resulting from static vision and active locomotion. For comparability, the same technological setup as in the first study was used. Nominal distances between 3.0 and 3.5 m allowed translation gains to range between 80% and 120% despite spatial restrictions of the tracking equipment. In contrast to our initial hypothesis, verbal estimates were found to be particularly accurate. While the results for visually guided walking tended towards overestimations, the remaining methods were affected by larger deviations. The influence of translation gains was confirmed for verbal estimates, but even more pronounced for methods that entailed actual or imagined walking. The latter finding suggests a higher sensitivity to motion-induced vestibular, proprioceptive, and visual information for certain methods.

Both studies highlight the potential to mitigate distance compression by active locomotion. Its effectiveness, however, may depend on the approach used to quantify estimates. Despite the identical technological setup and similar participant characteristics, the overall error was considerably smaller in the second experiment, indicating an influence of additional variables such as the number and order of trials that feature walking interaction.

B.2 Estimating Time to Contact

Schneider, S., Salloum, M., Gundel, K., & Boos, A. (Accepted). Estimating Time to Contact in Virtual Reality: Does Contrast Matter? In *Proceedings of the 21st Congress of the International Ergonomics Association (IEA 2021)*, Vancouver, Canada.

While distance compression is vividly discussed in VR-based research, pedestrian behavior may ultimately be more strongly affected by biases in the perception of alternative parameters. Anticipating the time until a moving object reaches the observer (time to contact, TTC), for instance, seems arguably more relevant to collision avoidance than merely assessing the distance to a vehicle at a given moment. Previous findings indicate that TTC is commonly underestimated in both virtual and non-virtual environments, and that some factors, such as variations in vehicle speed, affect estimates in both cases (Feldstein, 2019). There is, however, also reason to believe that impoverished visual perception, which is more likely to occur in virtual environments, may exacerbate biases (Cavallo & Laurent, 1988; Feldstein, 2019).

Contributing to the research on effects of limited physical fidelity, a simulator study (Schneider et al., Accepted) was performed to evaluate if changes in image contrast affect TTC estimates. Although low contrast has repeatedly been shown to influence speed perception (Horswill & Plooy, 2008; Snowden et al., 1998), it is yet unclear whether these effects generalize to TTC estimates in virtual traffic. In the current experiment, brief simulated scenes were presented via an Oculus Rift DK2 HMD, in which a single vehicle approached at a speed of 30, 40, or 50 km/h. At a distance of approximately 30 m, the screen turned black and participants were instructed to press a button when they expected the vehicle to pass them. Contrast was adjusted by a white overlay of 75%, 87.5%, or 100% transparency.

Differences in contrast had no significant effect on the expected TTC, whereas a previously observed tendency for higher estimates at higher speed was replicated. A closer examination of the mean values, however, suggests that the latter mainly concerned the increase from 30 to 40 km/h. Further factors, such as limited visual resolution and thus a limited capacity to discriminate movements at large distances, may alter this relationship at higher speed. Ratings of presence seemed relatively high but were unrelated to estimation accuracy. In addition, participants mostly thought they had underestimated the actual TTC, whereas the overall estimation error was small but positive. Considering the disagreement with previously observed changes in speed perception, the results indicate that physical fidelity may have distinct effects on the processing of related parameters such as speed and TTC.

B.3 Comparison of Simulator Types

Maruhn, P., Dietrich, A., Prasch, L., & Schneider, S. (2020). Analyzing Pedestrian Behavior in Augmented Reality — Proof of Concept. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, Atlanta, Georgia, USA, pp. 313 – 321. Available from https://doi.org/10.1109/VR46266.2020.00051

Pala, P., Cavallo, V., Dang, N. T., Granié, M.-A., Schneider, S., Maruhn, P., & Bengler, K. (2021). Analysis of Street-Crossing Behavior: Comparing a CAVE Simulator and a Head-Mounted Display among Younger and Older Adults. *Accident Analysis & Prevention*, 152(1), 106004. Available from https://doi.org/10.1016/j.aap.2021.106004

To account for prospective advancements in simulator technology, the experiment described in the third publication (Schneider et al., 2021) was extended by the development and evaluation of an Augmented Reality (AR) pedestrian simulator (Maruhn et al., 2020). A ZED Mini stereo camera was mounted on the HTC Vive Pro HMD to provide a video livestream of the test track environment. Virtual vehicles were superimposed to replicate the experimental conditions. In comparison to the original experiment, participants in AR experienced shorter trials, which excluded the time for the vehicles to return to their initial position. Nonetheless, the results appeared similar to the HMD simulator evaluated in the scope of the third article (Schneider et al., 2021): In comparison to the test track environment, acceptance rates declined in AR, with only 31% of gaps being accepted across all speeds and gap sizes. A similar delay was observed with regard to crossing initiation, whereas effects of vehicle speed on CIT were comparable in AR and on the test track. Again, participants reported that a collision would have been dangerous despite that fact that vehicles were purely virtual, but also stated that they found it harder than usual to decide when to cross. While a number of technological challenges are yet to be overcome, the parallels between the VR and AR approaches suggest that both setups may be used interchangeably, depending on the availability of an appropriate physical environment. Furthermore, since the two conditions featuring the same display type shared more similarities than those including the same visual database (i.e. the HMD and CAVE groups in Schneider et al., 2021), it appears that the former aspect may be more influential than the latter.

In the experiment outlined by Schneider et al. (2021), participants merely signaled their crossing intent. Pala et al. (2021), in contrast, evaluated differences between the two simulators described in Chapter 4.2 in the context of actual street crossing. In addition to a group of young adults (25 to 42 years), the participant sample included seniors between 64 and 81 years. In line with Schneider et al. (2021), gaps were accepted more frequently and crossing was initiated later for higher vehicle speed, which resulted in reduced safety margins. In the experiment reported by Pala et al. (2021), however, the proportion of accepted gaps was higher in the HMD than in the CAVE. Furthermore, crossings were initiated considerably earlier, with CITs in the HMD averaging more than a second before the preceding vehicle had passed. This finding is particularly noteworthy, since it is inconsistent with both the behavior observed in the CAVE and, for instance, a similar experimental scenario analyzed by Schneider and Li (2020). Similarly low CITs were obtained for different age groups and vehicle speeds. The analysis of interaction terms, however, revealed that differences between the simulators were larger in the younger group with respect to acceptance rates, whereas CIT differed more strongly for seniors, who started crossing earlier than young adults in the HMD, but later in the CAVE. In contrast to the CAVE, participants wearing the HMD appeared to slow down in the far lane, potentially for fear of colliding with the walls of the physical laboratory. Vehicle speed appeared more influential in the CAVE in the study by Pala et al. (2021), whereas comparable effects were observed for both simulators by Schneider et al. (2021).

In summary, these observations confirm once more that the relationship between simulator technology and the behavioral outcomes of pedestrian research depends on a range of different factors including the display type, the experimental task, and participant characteristics. Adding to the findings of the present thesis, the results may clarify important interactions and support the progress in VR-based pedestrian research.